以人为原点的
数据社会学分析方法

块数据 2.0

大数据时代的范式革命

大数据战略重点实验室◎著

块数据2.0 | 大数据引爆未来十大趋势
块数据场景应用八大预测

为什么这八种人必看这本书

"90后"、科技工作者、社会科学研究者、领导干部
新生代企业家、白领、社群领袖、创客



中信出版集团 · CHINACITICPRESS

VISIT...

LANZAROTE Caliente.com

以人为原点的
数据社会学分析方法

块数据2.0

大数据时代的范式革命

大数据战略重点实验室◎著

块数据2.0 | 大数据引爆未来十大趋势
块数据场景应用八大预测

为什么这八种人必看这本书

“90后”、科技工作者、社会科学研究者、领导干部
新生代企业家、白领、社群领袖、创客



中信出版集团 · CHINACITICPRESS

版权信息

书名:块数据2.0

作者:大数据战略重点实验室

ISBN:9787508661339

中信出版集团制作发行

版权所有•侵权必究

主编序

纵观全球，世界正处于新科技革命方兴未艾之时，我们正在迎来并开创一个创新的时代。在汹涌而来的创新浪潮中，顺之者昌，只有顺应规律、投身热潮，才能站在新时代的制高点上；逆之者亡，如果忽视潮流、置身事外，就会被时代所遗忘甚至抛弃；领之者强，唯有主动变革、引领潮流，才能真正立于不败之地。

放眼当前，大数据作为创新浪潮的最重要标志，其发展已经超出了绝大多数人的预期，从最开始的仅仅涉及产业领域到逐步渗透进人们的日常工作、生活乃至行为方式中，从少数人关注、少数人研究到成为整个社会高度关注的现象级词汇。立足实践，块数据作为大数据发展的高级阶段，凭借着对条数据的块上融合，依托着不同种类、领域、方面数据的自由流动和公开共享，依靠着以社会学为原点，对现有海量数据交和、融合、激活，以块数据的商用、民用和政用为探索途径，一大批的创新成果遍地开花，正在重构既有的经济社会体系，成为创新大潮中最闪亮的前沿与高端。可以讲，块数据时代的到来不可抗拒。

块数据是一个严肃的命题。虽然我们对它的认识还只是皮毛，但这些皮毛已经足以让我们心潮澎湃了。就像大数据是什么并不重要，重要的是大数据正在改变人们对世界的看法一样，块数据也是如此。如果仅仅把块数据当成一个概念、一种技术，或者说至多是推动产业变革或政府治理的新动能，那就过于简单了。块数据带来了一场新的科学革命，这场革命是以人为原点的数据社会学范式，核心是用数据技术而不是人的思维去分析人的行为、把握社会规律、预测人类未来。这是一场由科技引发的社会变革，它将改变我们的思维方式和社会生活方式，改变世界上物质与意识的构成，改变我们的世界观、价值观和方法论。我们不得不更多地关注它、研究它、把握它，因为我们每个人都身在其中。这也是我们必须对块数据肃然起敬，并且持续探寻的根本动因。

块数据是一个创新的课题。本书尝试从4个层面通过10个章节加以论述。第一个层面，包括第一章、第二章、第三章，提出块数据的定义，进一步阐释块数据的基本属性和本质，并指出块数据是大数据的解决方案；第二个层面，包括第四章、第五章，提出块数据概念模型，以

激活数据学为核心进一步阐释块数据的形成机制及其运行规律；第三个层面，包括第六章~第九章，提出块数据的全产业链、全服务链和全治理链，从块数据组织、块数据经济和块数据治理三个视角进一步阐释块数据在民用、商用和政用方面的应用价值；第四个层面，即第十章，提出块数据安全，既阐明块数据开放应用中的数据风险不可避免，又指出块数据安全防御的十大对策。我们试图探索并构建一个块数据理论的逻辑体系，以进一步揭示块数据的本质、规律和应用价值。

如果说大数据是对人类社会的一种预测，那么块数据则是在预测基础上对人类社会结构、经济功能、组织形态、生活方式和价值体系的一次重构。这种重构超越技术本身，是基于数据驱动的人机交互和人机合作，共同推动人类社会更高层次的螺旋式的持续进步。

块数据不是大数据的分支，更不是大数据的翻版，而是大数据发展的高级形态。块数据对人类的影响是全方位的。从条数据向块数据的汇聚、融合和裂变，将催生新的社会文明的崛起。数据力和数据关系的变化，深刻影响着生产力和生产关系的变革，而且正引发一场更为广泛的社会经济运动。块数据将会开启人类迈向共享社会的新征程，但人类也将毫无选择地进入无隐私时代和高风险社会。数据安全迫在眉睫，数权立法呼之欲出，数据伦理蓄势待发。

此外，块数据将激荡出令人心动的场景应用，DNA（脱氧核糖核酸）穿戴城市、云出行、长寿城市、惊艳旅游、泛在教育、360°全息媒体、消费4.0、超级链接组织等等。也许，不远的某一天，这些形态就会悄然而至……

我们在数据宇宙中努力探寻，发现世界，发现自我。有人问我，块数据是一门深奥的学问，我们看得懂吗？我的回答是：一种人，生下来就看得懂，因为他们是云脑思维；另一种人，可以看得懂，关键在于怎么看。从“懂”字的结构看，重要的东西总是埋在“草”下，唯有用“心”才能悟出。

这本书，为他们，也为我们而写：

云脑新人类——“90后”以及敢于挑战“90后”的人；科技工作者——魔法向左，科技向右；

社会科学研究者——人与社会正在被度量和数据化；

领导干部——不仅需要头脑风暴，更期待智能碰撞或人机交互；

新生代企业家——只使用最合适的人，不使用最聪明的人，最聪明的人永远在组织外部；

白领——培养一种有意义更有意思的数据观，尤其针对女性白领；

社群领袖——游刃于数据空间；

创客——无痛点，不创新。

大数据战略重点实验室就是一个块数据组织。它不仅仅是一个团队，更是一个平台。在这个平台上，外部的智慧与内部的创新相互汇聚、融合和激荡，每一个人都充满激情，每一个人都分享成功。大数据时代，什么样的人可以成功，那就是追赶未来的人。

是为序。

连玉明

大数据战略重点实验室主任

2016年4月1日于北京

绪论
大数据时代的范式革命

块数据是大数据时代真正到来的标志。如果说《块数据：大数据时代真正到来的标志》回答了块数据是什么的问题，那么《块数据2.0》就是进一步回答块数据为什么的问题。块数据带来了大数据时代的范式转移，颠覆传统的世界观、价值观和方法论，进而改变和形成新的知识体系、价值体系和生活方式，并将深刻影响人类政治、经济、文化和社会生活的方方面面。我们把这种改变和影响称为大数据时代的范式革命。

基于此，本书提出以下基本观点，并围绕“范式革命”这一主线展开讨论，试图进一步揭示块数据的本质和规律：

第一，以人为原点的数据社会学分析方法，更加强调用数据技术分析人的行为、把握人的规律、预测人的未来。

第二，激活数据学作为新的数据观和新的方法论，将实现对不确定性和不可预知性更加精准的预测。

第三，激活数据学的理论和应用，将催生块数据组织，促动数据力和数据关系的变化，深刻影响块数据经济和块数据治理，进而引爆新的社会文明。

对《块数据2.0》的研究过程，也是我们对块数据理论和方法进行探索的过程。这种探索所形成的一些新概念、新理论、新模型、新方法，既反映了本书的核心观点，也是我们对未来社会的前瞻预测研判。

（一）数据引力波揭示出从条数据到块数据的内在规律，从因果性到相关性的范式转移，以及人类社会从条时代迈向块时代的必然趋势。引力波的发现，标志着人类在破解宇宙奥秘的道路上又前进了一大步。而数据引力波的提出，犹如一把钥匙打开了多维世界的转换之门，让我们更加清晰地认识到大数据时代的运行规律和演变趋势，及其对未来社会发展带来的深刻影响。大数据时代的范式转移颠覆了传统的思维模式、资源配置模式和社会运行模式，跨界、无限、多维，以及物物相连、业业相连、人人相连带来的融合，打破了“条”的孤岛效应，使人、

物、数不再孤立，从条到块的开放共享成为可能。

（二）大数据时代是一个更加开放、更加复杂的巨系统，不确定性和不可预知性使一切坚固的东西都将可能烟消云散，激活数据学正成为大数据时代预测未来的新理论，自流程化成为应对不确定性的新方法。大数据时代，人类积累数据的能力远远超过处理数据的能力。垃圾数据泛滥、数据识别难度加大，以及数据采集、存储和使用方式发生重大变化，加剧了社会的不确定性和不可预知性。人类试图通过计算机、云计算或人工智能来解决这一难题，但并没有获得理想的答案，甚至产生了方向性的迷失。解决海量数据的困扰，应回归以人为原点的数据社会学的思维模式，以人机交互为突破，运用激活数据学的理论和方法，分析人的行为、把握人的规律、预测人的未来。这或许才是应对不确定性最精准、最有效的途径。

（三）数据作为国家基础性战略资源，正在形成以块数据价值链为核心的全产业链、全服务链和全治理链，发挥着引领全局、覆盖全面、贯穿始终的独特作用。以数据流引领技术流、物质流、资金流、人才流、服务流，并日益对全球经济运行机制、社会生活方式和国家治理能力产生重大影响，正在成为决定未来的潜在能力和关键因素。特别是数据作为一种创新驱动力，推动了“从1到N”向“从0到1”的转化。全产业链成为推动经济转型发展的新动力，全服务链成为创新公共服务的新模式，全治理链成为提升政府治理能力的新途径，块数据价值链的商用、民用和政用价值进一步凸显。

（四）块数据组织正演化为网络状组织的高级形态，数据人假设、数据力和数据关系的变化、平台领导力重塑，正孕育共享型组织的新范式。块数据组织通过外部资源强化自身战略地位，再平衡成为战略制高点。在块数据组织内部更为强调使用最合适的人，而不是最聪明的人。块数据组织更为强调平台化和外部性，通过外部力量打破组织界限，通过外包促进资源优化配置，数据力与数据关系的变化，转换成新的生产力和生产关系，重塑组织竞争优势。块数据组织通过联盟制的组织架构，实现联盟共存、联盟共创、联盟共治、联盟共担和联盟共享，最终形成共享型组织新范式。

（五）资源数据化促进了所有权与使用权分离，加速了效率与公平的高度统一。块数据经济正引爆一场更广泛的社会运动，人类开始从共享经济走向共享社会。资源一旦被数据化，必然带来再配置和再分配。

数据化的资源配置方式将带来所有权与使用权的分离，数据化的分配方式将加速效率与公平的高度统一。这种再配置和再分配，是一种更新中的社会体系，将形成新的社会经济模式，催生新经济的崛起。

（六）权力数据化和权力透明化正在解构和重构传统权力观，数据治理成为政府治理和国家治理的新趋势。贵阳全面推进“数据铁笼”工程建设就是对数据治理的最好诠释。“数据铁笼”是以权力运行和权力制约的信息化、数据化、自流程化和融合化为核心的自组织系统工程。“数据铁笼”工程的主体内容是构建开放共享的治理格局、规范透明的权力体系、跨界融合的平台支撑、持续改进的流程再造、精准有效的风险控制和多元治理的制度保障，关键是找风险、可评估、能预警、易处置、会防范。“数据铁笼”更加强化权力轨迹数据的归集、发掘和关联分析，更加强化权力风险预警和处置的数据支持，更加强化风险预测研判和智能防控。“数据铁笼”制约的不是一项权力，而是权力体系，包括科学确权、依法授权、廉洁用权、精准管权、多元督权的全过程，根本在于用好权、管住权，把权力关进数据的笼子里，让权力在阳光下运行，实现“人在干、云在算、天在看”。

（七）数据资源开放共享是块数据价值实现的重要基础和必要条件，政府数据开放成为开放数据的重中之重。共享难度大、垄断程度高、融合能力差、应用价值低、安全风险大是政府数据开放面临的主要问题。造成这种问题的深层次原因是条块分割、部门壁垒、技术障碍、创新应用差以及安全意识滞后、基础设施滞后、数据标准滞后和立法滞后。推动政府数据开放是开放数据的当务之急。以政府信息公开实现数据资源向社会开放，以公共数据资源交换促进跨部门数据资源的共享共用，以政府数据的契约式开放打通政府部门、企事业单位和社会组织的数据壁垒，有序推进政府、市场与社会对数据资源的合作开发和综合利用。

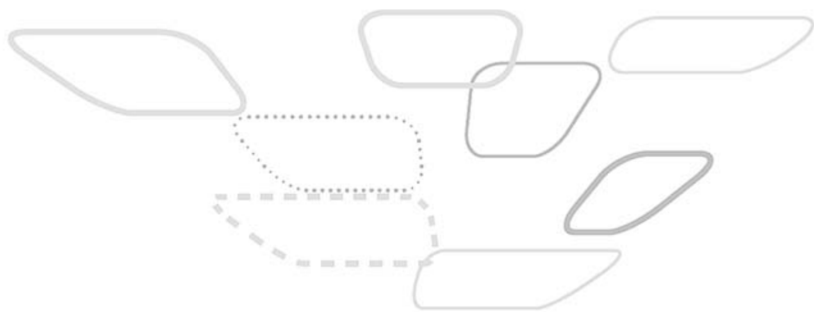
（八）数据滥用正成为一种不可逆转的社会常态，人类进入无隐私时代和高风险社会，数据安全面临新挑战。数据采集技术、数据存储技术、数据关联分析技术、数据激活技术和数据预测技术的广泛应用，使每个人或组织既成为数据生产者，又成为数据使用者。当数据的价值不断被挖掘和使用，甚至被当作商品待售或者已经出售的时候，数据滥用就必然异化成为一种社会常态，人类已无隐私可言。动辄百万级甚至千万级的数据安全事故所产生的威胁、侵害和影响，已远远超越个人和组

织范畴，正成为全球共同面临的系统性风险。大数据所引发的数据安全问题，并不在于技术本身，而是数据资源在开放、流通和应用中产生的各类风险。

（九）数据主权成为国家主权的新要素，保障数据资源权益，加快数权立法亟待提上议事日程。数据作为国家重要的基础性战略资源，网络空间数据主权正在成为国际竞争的新焦点。数据作为商用、民用和政用的核心要素，数据产权正在成为重塑竞争优势的新热点。建立完善国际合作机制，增强网络空间数据主权保护能力，发掘和释放数据资源的潜在价值，更好地发挥数据资源的战略作用。研究推动数据资源权益立法，加快建立数据市场交易标准体系，建立健全数据资源交易机制和定价机制，鼓励市场主体进行数据交换和交易，规范交易行为，促进数据资源开放、流通和应用。

（十）利他主义的数据文化催生新的社会文明，一场新的伦理革命蓄势待发。大数据不仅是新的科技革命和产业变革的引爆点，更是一种新的世界观、新的价值观和新的方法论。从条数据到块数据的融合，从条时代向块时代的迈进，整个人类社会的思维模式和行为范式将产生根本性、颠覆性变革。块数据倡导开放、融合、共享的价值理念，利他主义将成为新数据时代的主流文化，并孕育着一种新的社会文明。社会文明的进步从来都不是一蹴而就的，新的社会文明的出现必然伴随着文明的冲突和伦理的重构，虚拟与现实、利己与利他、规范与自由、封闭与开放、权威与民主之间的冲突和妥协必将引发一场新的伦理革命。

第一章
重新定义大数据



从人类思维范式的进程看，每个阶段的认知体系和由此产生的思想工具是不同的。第一个阶段，知识就是力量，知识是人脑思维的产物。第二个阶段，信息就是能量，信息是电脑技术的产物。第三个阶段，数据就是变量，无边界的数据聚合是人脑和电脑的思维范式无法完成的，必须是人、智能机器和云计算的融合，是一种云脑思维。换句话说，人类的思维范式分为三个阶段，即人脑时代、电脑时代和云脑时代。

大数据正在引发人类社会变革。它已经不是一个新概念，而是正在到来的新时代，是一种新的信息技术、新的服务业态、新的研究范畴和新的思维模式。数据成为一种多维的无限变量。数据质点由于价值关联产生数据引力，进而形成数据引力波，成为大数据超级能量的源泉。如同爱因斯坦提出引力波一样，数据引力波的提出推动了多维世界的价值转换，从因果性到相关性的范式转移，标志人类开始从条时代迈向块时代。

第一节 知识、信息和数据

（一）知识、信息和数据的双向演进

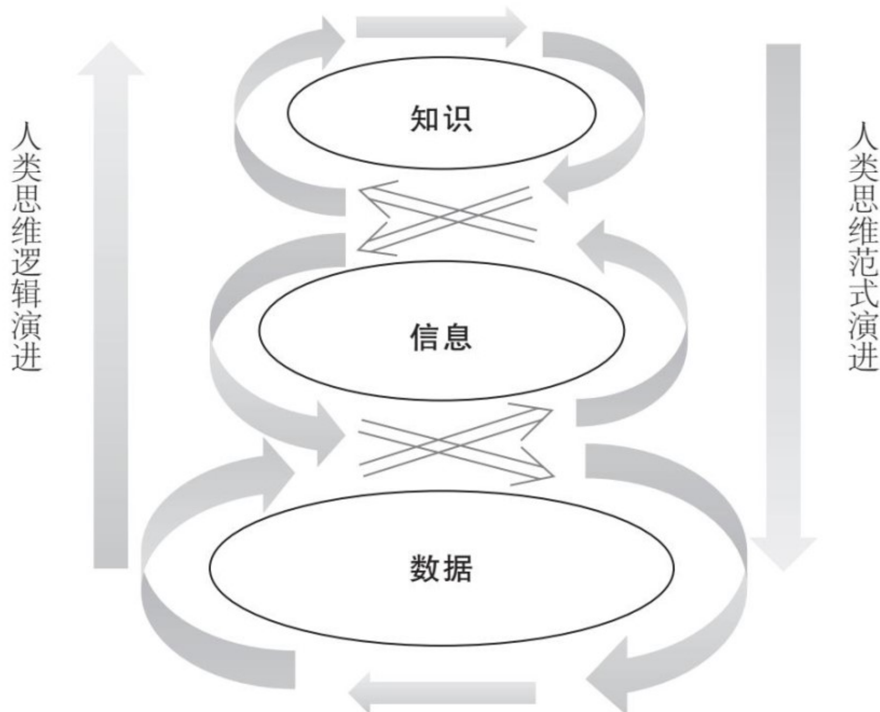


图1–1 知识、信息和数据的双向演进

从人类思维逻辑的进程看，数据、信息、知识是螺旋上升的循环周期。“数据——信息——知识”是处于一个平面上的三元关系，分别从语法、语意以及效用三个层面反映了人脑思维逻辑的过程。首先是初级过程思维阶段，这一阶段主要是对传统的数据（未经组织的数字、词语、声音、图像等）的记录，它可以直接来自测量仪器的实时记录，也可以来自人的认识。其次是次级过程思维阶段，这一阶段是信息的创造阶段，通过人脑次级过程思维活动，实现对原始数据的筛选、加工、创造，进而产生有意义的数据，也就是信息。最后是人脑的高级过程思维阶段，在这一阶段，人脑运用信息系统，对信息和相关的知识进行规律性、本质性和系统性的思维，对获取或积累的信息进行系统化的提炼、研究和分析，进而形成知识。作为人脑思维过程的三个阶段，数据、信息、知识是螺旋上升的循环周期。人们运用信息系统，对信息和相关的知识进行规律性、本质性和系统性的思维活动，创造新的知识。之后，新的知识又开辟了需要进一步认识的对象领域，然后使人们补充获取新的数据和信息，进入新一轮的循环上升周期。

从人类思维范式的演进看，知识、信息、数据呈现出一种螺旋上升

的循环周期。思维范式是指立足于一种世界观、认知体系、信念等而形成的固有的、稳定的、反复使用的具有范例特点的思维规范、模型或模式。思维范式是在人类社会发展过程中形成的，与其所处的历史阶段密切相关。思维范式与时代的发展是辨证互动的关系，人类思维范式的每一次发展演进对于时代的影响和作用都是巨大的，思维范式的创新是人类社会进步的重要动力。纵观人类思维范式的发展历程，可以分为人脑时代、电脑时代和云脑时代三个阶段，与之相对应的就是知识、信息和数据不断演进的过程。在人脑阶段，人类思维的界限就是人脑认知的边界，知识在推动社会进步的同时也在不断扩大人类活动的范围，推动知识的边界不断外延。电脑时代的到来是以信息技术革命为标志的，计算机特别是互联网的出现，产生了大量不能被人类所感知的信息，使得人类思维的范式已经不再局限于人脑思维，人类进入电脑时代。第三阶段是云脑时代。数据超出了人脑思维和电脑思维所能承载的范畴，需要借助人、智能机器和云计算的融合形成云脑思维并将人类带入云脑时代。人类思维范式的演进过程，是人类思维由繁到简的逆向思维过程，既是推动人类社会不断进步的力量，也是人类社会不断发展的结果。

无论是在人脑时代、电脑时代还是云脑时代，人类思维逻辑和人类思维范式都是同时存在的，数据、信息和知识都是其中最基础性的资源，三者相互交叉、相互促进，共同推动人类发展。

（二）知识就是力量

知识是人类在认识和改造客观世界中对实践或经验的记录和整理，是人的思维的产物，是人类智慧的最基本因素，其主要表现形式为数字、符号、文字等。数字产生于原始社会，是人类初始性的知识。符号是人类早期原始的表述手段，具有表意的功能，标志着人类从表象认识向兼具形式和意义的深层次认知过渡。文字是现阶段人类文明中最稳固、最先进、最复杂的知识表现形态，是人类社会进入文明时代的重要标志。文字的出现引发了人类知识在时间和空间两个维度的巨大飞跃。

英国著名的科学家、哲学家弗朗西斯·培根提出“知识就是力量”的论断，开辟了人类把知识作为一种重要资源的先河。

知识是人的思维产生和更新的内生力。思维是一种高级认识活动，知识是思维活动的基本构件，为人的思维活动提供原始素材，并规定、影响着思维活动的方式和方法。注思维具有自我更新能力，利用人脑

中已掌握的知识和产生的经验，通过联想、想象、直觉、灵感、发散等方式，发现尚未知晓的或者尚未出现的事物，推动新知识的产生，进而实现思维的更新。知识激发人的思维能力，是人类智慧最集中的体现，使得人类突破自然极限，在一切领域开创新局面成为可能。

知识是推动人类社会发展的驱动力。社会文明的发展进步皆是建基于知识，没有知识便没有社会进步。知识向生产力的转化，关键在于运用知识的能力，而非任何知识的简单集合。知识的不断更新能够提供人类征服和改造自然的新方法、新工具和新途径，知识的介入可以加快科学技术转化为生产力的进程。在时代的更替过程中，知识始终是推动人类社会突飞猛进的重要力量。

德国哲学家康德认为，“自在之物”与“现象”（即“此岸”与“彼岸”）之间存在着原则上的界限，是人类认识无法逾越的鸿沟。人们往往只能认识“自在之物”的表象，而不能透过现象去认识“自在之物”的本质。也就是说，知识本身是无界的，但人类对知识的认识是有边界的。

一方面，人脑认识和承载知识的能力是有限的。人脑作为生命有机体的一部分，是有生命周期的。一旦生命有机体死亡，人脑及其所存储的知识也一并消亡，正如《庄子·养生主》中所说，“吾生也有涯，而知也无涯，以有涯随无涯，殆已”。人类生命的短暂性决定了人对知识的认识是有限的。同时，人类对知识的承载和处理能力也是有限的。认知负荷理论认为，人类的工作记忆系统同时加工新信息的容量是有限的，为了使加工得以顺利进行，当前进入工作记忆的信息量不能超过工作记忆的容量[注]，而且人类的大脑也没有足够的空间能够装下所有知识。知识超载让人类大脑容易产生困惑，难以提取有效知识并做出理性判断。

另一方面，人类的认知能力是有限的。人类所获得的知识只限于能够被感知的显性知识，很多隐性知识则超出了人类认知的范畴，特别是对于空间上遥远的地方和时间上久远的过去，人的所知更少。由此，显性和隐性之间就存在着人类认知的边界，这种边界处在不断运动和发展中，在时间上没有开始和终了，在空间上没有边界和尽头。

信息化打破了知识的边界。在信息化时代，以计算机技术为基础，电脑代替人脑进行信息的记录、筛选、传播，打破了人脑对知识记忆、存储、处理的局限性。同时，信息的积累与交换、分析与运用，改变了人类获取知识的方式，突破了获取知识的时空限制，人类采集信息进而

获取知识的能力显著增强，并产生了前所未有的知识量，加快了知识转化为生产力的速度。

（三）信息就是能量

信息论奠基者香农认为，信息是用来消除不确定性的东西，是不定性减少的量，是两次不定性之差。[注]随着人类社会的发展，信息量不断增多、传递速度不断加快、共享范围不断扩大，离开信息的传播、交流和融合，任何发明创造都难以实现，而信息的增加、积累、再生，必然要依靠信息载体的进步和革命。

20世纪中叶，以计算机和互联网为标志的第五次信息革命，实现了计算机技术和通信技术的融合，克服了时间和空间的阻碍。信息采集、传播的速度和规模达到空前水平，实现了全球信息的交互与共享，信息量的增长速度超越人类的理解速度，开启了一个信息大爆炸的时代。据英国学者詹姆斯·马丁统计，人类知识的倍增周期，在19世纪为50年，20世纪前半叶为10年左右，到了70年代，缩短为5年，80年代末几乎已到了每三年翻一番的程度，全球信息量呈指数级增长，并且增长速度不断提高。

随着信息载体的不断更新，信息传递的效率也在不断提高。从书信到电话，信息是1对1的传播模式，传递效率值是1。电视广播的发明使得同一信息能够让N个人接收，传递效率从1提高到N。互联网的出现，使每一个人既是信息的接收者，也是信息的传播者，信息传播的节点可以无限增多，信息通路可以无限增多。鲍勃·麦特卡尔夫提出的麦特卡尔夫定律认为，网络价值同网络用户数量的平方成正比，即N个连接能创造N^2的效益，N个人能看到N个人的信息，信息传递的效率是$N \times N = N^2$。

与以往不同，在互联网时代，信息的传播、存储、处理主要依托计算机完成，信息超出了人脑所能认知的范畴。某种意义上可以说，人类认知边界内的即为知识，认知边界外的即为信息，这个边界即为人脑与电脑的界限。但信息是一种尚未被人类所完全认知的客观存在，依赖能量为人类所感知，不同信息系统通过交换、演变、融合、转化，实现能量的聚合，并在一定条件下，这种能量能够被释放出来，对人类社会产生重大影响。马克·布尔金在《信息论》中提出，信息支配着社会，渗透到每个生活领域。[注]信息能量的释放总是伴随着新一轮的积聚，在积

聚中转化，在转化中释放更大能量。

信息数据化是信息释放能量的手段。“数据化”不是“数字化”。数字化仅仅是把模拟数据转换成0和1表示的二进制编码，实现计算机的可识别、可读取和可处理。数据化则是指一种把现象转变为新型可分析数据的量化过程，包括数据的采集和数据的处理，也就是对某些事物的描述，数据可以记录、分析和重组。借助计算机技术、通信技术和高密度存储技术，信息可以更快、更准地转变为信息资源。而且，信息被数据化之后，就具有通用性、开放性、标准化和高度整合性的特点，信息无限量的积累和跨时空、跨行业的传递，有利于释放信息潜在的巨大价值和能量。

信息数据化标志着人类认识的一个根本性转变。通过信息数据化，能够全面采集和计算客观存在的信息，并对其进行加工处理。在数据化时代，我们将改变把世界看成自然社会和人类社会的观念，从数据的内核上来认识世界，将世界看成各类信息的集合。这为人类认识世界、改造世界提供了一种全新的视角，是一种可以渗透到所有生活领域的世界观。

（四）数据就是变量

从知识到信息再到数据，是人类文明进步的历程。数据既是自变量也是因变量，在自身变化的同时，也会引发外部世界的变化，并且数据的自变与因变的过程是同时发生作用的。在大数据时代，数据是一种重要的基础性战略资源，成为引发世界变革的重要因素。

数据是一种客观存在，是对现实世界的数据映射。数据自身处于运动变化之中，是信息通信技术按照自身发展逻辑，从提高生产效率向更高级智能阶段的自然生长。从数量上来看，大数据不仅是庞大的，而且是不断变化的。国际数据公司发布的《数字宇宙》报告，对全球的数据存量和增长趋势进行了定量评估。报告显示，全球IP（网络之间互连的协议）流量达到1EB（艾字节），在2001年需要1年，在2013年仅需1天，到2016年则仅需半天，至2020年，数字宇宙规模将增长10倍，每年产生的数据量由当前的4.4万亿GB（吉字节），增长至44万亿GB。从性质上来看，跨界、跨领域的关联和重组是数据自身发展的天性，它能够打破时空的界限进行快速的流转和聚合，并把同一类型、同一领域的数据聚集成类，相互作用，并形成更高层级的、跨领域的持续集聚，进

而在新的条件下形成新的价值。数据的多种形式、多种来源以及数据之间的多种复杂的联系，都使数据世界变得更加神秘也更加激动人心。

数据成为改变世界的战略资源。麦肯锡公司的报告指出，“数据，已经渗透到当今每一个行业和业务职能领域，成为重要的生产因素。人们对于海量数据的挖掘和运用，预示着新一波生产率增长和消费者盈余浪潮的到来”。通过技术的创新与发展，以及全面感知、采集、分析、共享，数据为人们提供了一种全新的看待世界的方法。当我们运用数据来决策，并对事物的发展趋势进行预判时，世界变得可知和可预测，数据就成为一种不确定性对抗确定性的变量。当前，国家间的竞争正从对资本、土地、人口、资源或能源的争夺转向对数据的占有。在大数据时代，作为一种新的战略资源，数据改变了资本和土地等传统生产要素在经济发展中的权重，成为与高端人才、稀有能源一样重要的提升国家竞争力的战略制高点。大数据颠覆性地改变经济形态、国际安全态势、国家治理和资源配置模式，引发了经济社会的巨大变革。[注]以数据流引领技术流、物质流、资金流、人才流，持续激发新模式，形成新技术，催生新经济，将成为推动人类社会生产和生活方式发生根本性变革的核心力量。

（五）数据引力波

引力是人类最早定量认识的相互作用，它无处不在，主导了宇宙和星系。在经典物理学中，引力与强力、电磁力、弱力一同被认为是宇宙中四大基本力之一，是最弱的力，但也是起决定作用的力。引力的大小与质量成正比，与距离的平方成反比。在量子力学中，引力被认为是由于两个粒子交换引力子导致的。爱因斯坦广义相对论认为，引力起源于质量对空间的扭曲，是空间——时间弯曲导致的几何效应，在任何有质量的物体附近，空间的结构都会受到扭曲，但这样的扭曲并不总是围绕在物体的附近。按照广义相对论的界定，质量导致时空结构弯曲，进而产生引力，而产生引力的空间是不平坦的、运动的、多维的和复合叠加的，它没有长度、时间、密度等有形物质属性。

在对引力进行了一个全新维度的界定之后，爱因斯坦也意识到空间的弯曲形变能在宇宙中进行传播，类似于地震波在地壳里传播，这种传播是通过引力波实现的。爱因斯坦认为，引力波是时空曲率的扰动以行进波的形式向外传递，并通过引力辐射的形式传输能量。引力波在物体加速运动过程中产生，即物质的分布发生改变时，比如恒星爆炸、黑洞

碰撞，都会产生引力波，会引起时空的伸缩、影响时空的结构。引力波发挥作用的形式与流体力学中的重力波类似，重力波是液体介质内或两种介质面间的一种波，当液体表面或内部液团由于密度差异离开原来位置，在重力和浮力的综合作用下，液团会处于上下振动以达到平衡的状态，即产生波动。引力波的本质就是时空曲率的波动，也可以称之为时空的“涟漪”，就如同石头被丢进水里产生的波纹一样，这种波纹既会层层扩散，又能相互影响，循环往复。通过引力波释放出的能量超过了整个宇宙中所有星光的能量，引力波所携带的不仅有能量，还有关于它们产生过程中的信息。

在大数据时代，数据以其庞大的数量和特有的运行模式构成了与传统社会不同的新的时空——数据宇宙。数据质点是数据宇宙中具有“数据质量”的数据单元(注)，由数据空间中一组有特定关系的数据元素（点）组成。一般来说，这种关系是指数据元素在数据空间中的几何相邻关系。在数据质点的作用下，数据空间会发生不规则、无边界、多维度和复合叠加的运动，进而产生数据引力，这与广义相对论中引力产生的时空扭曲类似。不同于物理世界中的引力，数据引力是一种标量，没有方向性，表示的是数据之间的价值关联性。

广义相对论认为加速运动的质量会产生引力波。同理，数据质点的快速流动也会产生数据引力波。任何数据之间都存在数据引力波，引力波的大小取决于数据质点价值关联性的强弱，即数据质点间的价值关联性越强，则数据引力波越大；价值关联性越弱，则数据引力波越小。数据引力波的特点是具有激活性、穿透性和不可预知性。数据引力波是一种隐性波，只有在激活的状态下，数据的价值才能显现出来。数据的激活分为自激活和他激活，都是在数据引力场中完成的。

数据本身具有不同的类型、性质和维度，同一类型、性质和维度的数据质点在数据引力的相互作用下，形成一个充满数据的场域，即数据引力场，数据引力场的形成是数据量的积累过程。数据的多维性决定了任何一个数据都可以同时存在于多个引力场中，并产生不同的数据引力波。同时，数据与数据引力场、引力场与引力场之间都存在数据引力波。通过数据引力波，大量的数据或者数据引力场快速碰撞、相互融合，实现聚合、裂变，产生新的价值关联，并释放巨大的能量。这种能量远超数据本身的价值总和，甚至超出人类想象的范围，数据引力波将这种巨大的能量传送到无限远处。在这一过程中，原有数据引力场的平

衡力被打破，借助数据引力，大量的数据重新聚合，重组再造新的数据引力场，实现新一轮的数聚和裂变。这是一种解构到重构的过程，也是量变到质变的过程。依托数据的解构和重构，数据宇宙实现膨胀和压缩。例如，黑客数据进入特定的数据空间，扰乱原本数据引力场中数据与数据之间或者数据引力场之间统一的规则秩序，这一过程就是数据与数据引力场相互碰撞并产生作用的过程，进入引力场的黑客数据破坏了原有数据空间的平衡力，成为“不速之客”。

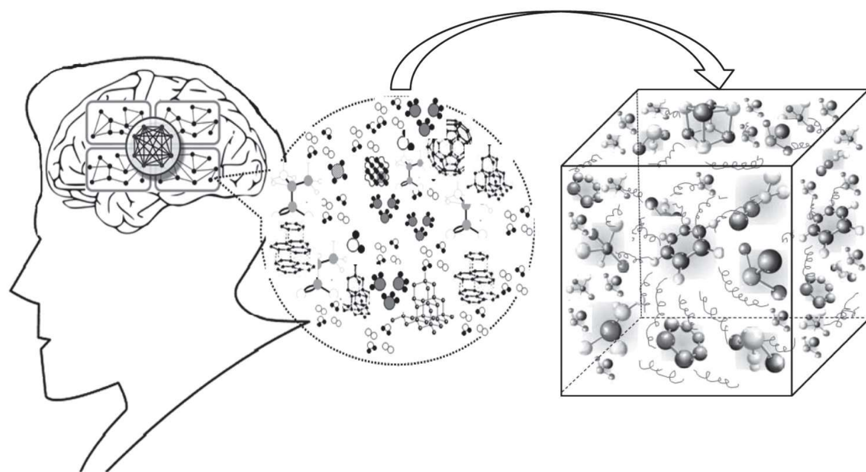


图1–2 数据引力场断面示意图

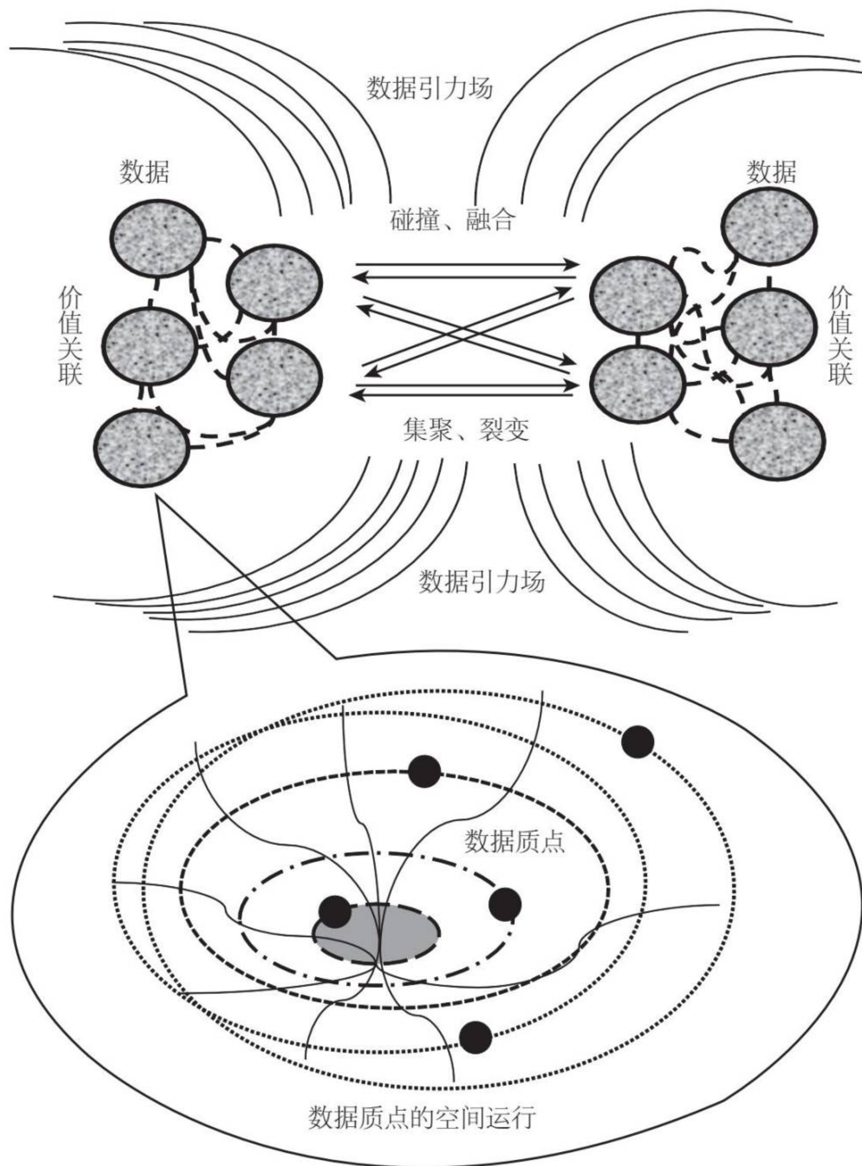


图1–3 数据引力波的形成过程图

如今，引力波的证实成为科学史上最重要的一刻。数据引力波也将成为人类认识数据宇宙的钥匙。它不仅会改变人类既有的经验积累，推动一系列基础理论学科实现质的飞跃，也会改变人类既有的思维范式，形成全新的数据世界观，引领人类社会真正进入云脑时代。正是通过数据引力波所引发的数据在多维空间的集聚和裂变，为大数据打破“条”的限制，实现“块”的聚合提供了可能，为数据在多维空间内实现自激活、自流程、自适应、自组织创造了原始动力。

第二节 大数据是什么

（一）大数据溯源

20世纪60年代，伴随着数据库的出现，人类采集、存储数据的能力迅猛提高，数据的总量不断增加，而且增加的速度不断加快，形成海量数据。在数据量快速积累的同时，数据的内涵也在不断扩大，除了测量和计算之外，它又具有了记录的功能。数据的记录功能与信息类似，但范畴要大于信息。比如，一封邮件虽然包含很多条信息，但是从技术的角度看可能还是“一个数据”。海量数据的出现，使传统的计算科学范式逐渐被抛弃，摩尔定律逐步成为数据处理的重要技术。从数据类型上看，这一阶段的数据还是以结构化数据为主。但是，伴随着社会媒体、移动终端和传感设备等新渠道和新技术的大量出现与应用，包括文档、图片、XML（可扩展标记语言）、HTML（超文本标记语言）、图像和音频/视频信息等在内的非结构化数据和复杂数据产生，人类迎来了大数据时代。

1890 年
- 由赫尔曼·霍勒瑞斯发明的可以由机器处理的穿孔卡片，突破了传统人口普查的困难，该设备让美国用一年时间就完成了原本耗时 8 年的人口普查活动，在全球范围内引发了数据处理的新纪元。

1943 年
- 英国为了破译“二战”期间的纳粹密码，开发了系列开创性的能进行大规模数据处理的机器，并使用了第一台可编程的电子计算机进行运算。它以每秒钟 5 000 字符的速度读取纸，破译德田部队前方阵地的信息，帮助盟军成功登陆了诺曼底。

1989 年
- 英国计算机科学家蒂姆·伯纳斯·李开创了一个叫作“万维网”的超文本系统，在全球范围内利用互联网实现信息共享。

1997 年
- 美国宇航局研究员迈克尔·考克斯和大卫·埃尔斯沃斯首次使用“大数据”这一术语来描述 20 世纪 90 年代面临的数据挑战。数据集之大，通常超出了主存储器、本地磁盘的存储能力，甚至远程磁盘的承载能力，故而称之为“大数据问题”。

2007 年
- “大数据”一词开始在技术圈内出现。《连线》杂志发表文章阐述了数据泛滥带来的机遇和挑战，称大数据是“Petabyte（拍字节）时代”的开端。

2008 年
- 计算社区联盟作为最早提出大数据概念的机构，发表《大数据计算：在商务、科学和社会领域创建革命性突破》白皮书，提出“大数据真正重要的是新用途和新见解，而非数据本身”。

2010 年
- 肯尼斯·库克尔在《经济学人》上发表了大数据专题报告:《数据，无所不在的数据》。他在报告中提到:“世界上有着无法想象的巨量数字信息，并以极快的速度增长。”库克尔因此成为最早洞见大数据时代趋势的数据科学家之一。

2011 年
- IBM（国际商用机器公司）的“沃森”超级计算机每秒可扫描并分析 4TB（4 太字节，约 2 亿页文字量）的数据量，并在美国著名智力竞赛电视节目《危险边缘》(Jeopardy）上击败两名人类选手而夺冠。《纽约时报》将这一刻称为“大数据计算的胜利”。

2012 年
- 瑞士达沃斯召开的世界经济论坛上，大数据是主题之一，会上发布的报告《大数据，大影响》宣称，数据已经成为一种新的经济资产类别，就像货币或黄金一样。

2014 年
- 世界经济论坛以“大数据的回报与风险”为主题发布了《全球信息技术报告》(第 13 版)，美国白宫发布了 2014 年全球“大数据”白皮书的研究报告《大数据：抓住机遇、保存价值》。

图1–4 大数据概念形成初期的重要事件

20世纪90年代兴起的复杂性科学(注)，为人类提供了有机自然观，整体全局、关联、演化的复杂性思维方式和新的科学理论方法，也为大数据提供了理论基础。以互联网、云计算和人工智能为代表的新兴技术为大数据的出现提供了技术条件。同时，谷歌、脸谱网、亚马逊、百度、阿里巴巴、腾讯等大数据相关企业的崛起推动数据产业快速发展，

为大数据发展提供了社会基础。基于此，著名未来学家托夫勒在《第三次浪潮》中正式提出大数据的概念，并将之称颂为“第三次浪潮的华彩乐章”。《自然》杂志也于2008年推出了“大数据”的封面专栏，讲述了数据在数学、物理、生物、工程及社会经济等多学科领域扮演的重要角色。

（二）大数据概念辨析

关于大数据的定义最早是来自企业而非学术性机构，这与大数据最早源于企业有关。即使到目前，对于大数据概念的讨论也一直没有停止，学术界、产业界、政府机构都从自身的研究领域和立场出发，在性质特点、要素构成、技术系统、应用范围、价值来源等方面对大数据的内涵和外延进行了不同的界定。

技术分析角度。这一类定义重点关注的是对海量、复杂数据进行分析处理，从而获得信息和知识的技术手段。例如，麦肯锡就认为，大数据是大小超出常规数据库工具的获取、存储、管理和分析能力的数据集[注]，也指无法采用传统流程、工具处理或分析的信息，迫使用户采用非传统处理方法的数据集，数量级不一定要超过特定的数据存储容量值。而维基百科认为，大数据是指无法在合理时间范围内用常规软件工具进行捕捉、管理和处理的数据集合。[注]数据集成软件商纳斯达克则认为，大数据包括海量数据和复杂数据类型，其规模超过传统数据库系统进行管理和处理的能力。[注]

大数据应用价值角度。这一类定义强调的是大数据的应用，关注的是从数据中获取有价值的信息和知识，最终目的是建立商业方面的竞争优势甚至是创新商业模式。比如，高德纳咨询公司认为，大数据是需要新处理模式才能具有更强的决策力、洞察力和流程优化能力的海量、高增长率和多样化的信息资产。维克托·迈尔–舍恩伯格认为，大数据时代的来临使得人类第一次有机会和条件在非常多的领域和非常深入的层次获得和使用全面数据、完整数据和系统数据，深入探索现实世界的规律，获取过去不可能获取的知识，得到过去无法企及的商机。[注]哈佛大学访问学者徐晋在《大数据经济学》中指出，大数据是指存在价值关联的海量数据。大数据的本质是社会经济的离散化解构与全息化重构，表现为行业间海量数据的关系从量变到质变的转变（深度挖掘）。赵国栋、易欢欢等在《大数据时代的历史机遇》一书中也指出，大数据是在

多样的或者大量的数据中迅速获取信息的能力。

大数据自身特征角度。这一类定义是从大数据本身性质和特点对大数据进行界定。其中又可以分为以下两个角度。一是从量的角度，比如百度百科认为，大数据或称巨量资料，指的是所涉及的资料量规模巨大到无法通过目前主流软件工具，在合理时间内达到撷取、管理、处理并整理成为有助于企业经营决策的资讯。大数据科学家约翰·劳瑟认为，大数据就是任何超过了一台计算机处理能力的庞大数据量。二是从性质的角度，高德纳咨询公司分析师道格·兰尼首次提出大数据的3V特征，即高速增长的数据体量（Volume），高速进出的数据运动（Velocity），高度异质的数据种类（Variety）。在此基础上，麦肯锡公司在《大数据：创新、竞争和生产力的下一个前沿》中提出了大数据具有4V的特征，即：数据容量大（Volume）、数据类型繁多（Variety）、商业价值高（Value）、处理速度快（Velocity）。

大数据对社会发展影响角度。这一类定义强调大数据对人类社会生产生活方式、思维范式等产生的重大影响，认为大数据开启了人类发展的新阶段，并且认为这种范式的影响是持久而深远的。比如，维克托·迈尔–舍恩伯格教授指出，“大数据”所代表的是当今社会所独有的一种新型的能力——以一种前所未有的方式，通过对海量数据进行分析，获得有巨大价值的产品和服务，或深刻的洞见。维克托·迈尔–舍恩伯格在《大数据时代》中前瞻性地指出，大数据带来的信息风暴正在变革我们的生活、工作和思维，大数据开启了一次重大的时代转型。中国工程院院士李国杰认为，理解大数据需要上升到文化和认识论的高度。数据文化的本质是尊重客观世界的实事求是精神，重视数据就是强调用事实说话、按理性思维的科学精神。(注)

（三）大数据定义

大数据时代，“数据驱动”成为全球大势。大数据的发展对经济社会发展和人类的思维观念带来革命性的影响，成为美国、英国、欧盟等许多发达国家和地区的重要发展战略。

中国数据资源丰富，应用市场前景广阔。互联网、移动互联网用户规模居全球第一，涌现出百度、阿里巴巴、腾讯这样顶尖的互联网创新型标杆企业。大数据正日益对中国的生产、流通、分配、消费活动以及经济运行机制、社会生活方式和国家治理能力产生重要影响。2015年8

月，《促进大数据发展行动纲要》的颁布，成为推动中国大数据发展的重要顶层设计和战略部署。党的十八届五中全会提出实施“国家大数据战略”，标志着大数据战略正式上升为国家战略。2016年3月，《中华人民共和国国民经济和社会发展第十三个五年规划纲要》，明确提出要把大数据作为国家的基础性战略资源。《促进大数据发展行动纲要》对大数据进行了全新界定，即“大数据是以容量大、类型多、存取速度快、应用价值高为主要特征的数据集合，正快速发展为对数量巨大、来源分散、格式多样的数据进行采集、存储和关联分析，从中发现新知识、创造新价值、提升新能力的新一代信息技术和服务业态”。这是国家层面对大数据最具权威的官方解读。这一新的定义，蕴含着大数据时代的三个基本特征，即新模式、新技术、新业态。科学认识这些特征可以帮助我们去探寻大数据带来的变化以及这些变化是如何发生的，这正是发现大数据的本质的过程。

新模式：核心是新的思维范式。大数据不仅是一场技术革命，更是一场思维的革命。大数据思维范式的关键转变在于从人脑思维到电脑思维再到云脑思维的转变，这种思维具有以下特点。一是总体性，伴随着数据在采集、存储、分析等相关技术上的突破，对于数据的获取实现了从样本数据到全体数据的转变。二是容错性，精确性是小数据时代的产物，当数据量无限大时，绝对的精准不再是数据追求的主要目标。三是相关性，也就是人们只需知道“是什么”，而不用知道“为什么”。四是智能性，智能是大数据时代的显著特征，思维方式从自然思维向智能思维转变，不断提升机器设备或系统设置的社会计算能力和智能化水平，从而获得具有洞察力和新价值的数据，甚至类似于人类的智能。

新技术：核心是新的信息技术。大数据本身是什么并不重要，重要的是大数据背后蕴含的价值所带来的影响。大数据具有“容量大、类型多、存取速度快、应用价值高”和“数量巨大、来源分散、格式多样”的特征，大数据的价值在于应用，必须依靠全新的处理方式，即新的数据采集技术、数据存储技术和数据关联分析技术，从根本上解决“数据从哪里来、数据放在哪里、数据如何使用”这三大问题，实现通过数据发现新知识、创造新价值、提升新能力的目标。

新业态：核心是新的服务业态。大数据带来社会生产要素的开放共享、集约整合、协同开发和高效使用，改变了传统的生产方式和经济运行机制，持续激发商业模式创新，不断催生新业态。这种新业态以

新的服务业态为核心，通过大数据驱动传统服务模式或商业模式的再创新。大数据在重构未来经济格局的同时，也将对传统的社会关系带来重构。

第三节 从大数据到块数据

（一）海量数据的悖论

海量数据在推进人类社会发展的同时，也带来了新的问题和困扰，我们把这种问题和困扰称为“海量数据的悖论”。

海量数据带来数据量激增的同时，数据分析的不确定性也在增长。维克托·迈尔–舍恩伯格认为，“就是预测，不是要教机器像人一样思考，而是要把数学计算运用到海量数据上，来预测事情发生的可能性”。在数据“匮乏”的时代，数据的采集、搜索等技术手段的落后导致人类由于可获得数据的有限性，难以对事物做出准确的判断和预测，如同处在黑暗中，辨不清方向。[注]进入大数据时代，数据短缺转变为数据过剩。信息爆炸与数据爆炸，带来海量信息和垃圾数据泛滥的困境。客观上讲，在数据稀缺的时代人们希望数据多多益善，而在信息爆炸和数据爆炸的时代人们则是希望数据越精准越好、越有序越好。人类获得优质数据的难度较以往不是降低而是提高了。海量数据的复杂性不仅体现在数据的体量上，更体现在多源异构、多实体和多空间之间的交互动态性[注]，以至于我们不能用传统的方法去描述与度量这些数据。数据分析者、研究者与决策者再一次陷入结构性短缺、集团化占有、分散化使用和有限性分享等新的数据困惑之中[注]，这直接影响了大数据预测的准确性和客观性。

（二）数据处理技术及其趋势

海量数据对数据的处理技术提出了更高的要求。当今世界正面临着由技术瓶颈突破带来的全行业升级，云计算、人工智能技术在推动人类社会全面进入大数据时代的同时，在数据处理方面也具有一定的局限性，以人的行为为原点，以社会学为基础，进行多维度数据分析，将成为人类社会大数据处理技术的发展趋势。

云计算数据处理能力提升是未来大数据处理的技术路径。在云计算

出现之前，摩尔定律一直主导着数据处理技术的发展。摩尔定律受限于芯片性能的提高，而云计算恰恰能够很好地解决这一问题。云计算为大数据提供了按需扩展的计算和存储资源，云计算技术的一个核心理念就是通过不断提高自身的处理水平来减轻用户负担，从而使用户终端简化成单纯的输入输出设备，并能够按需享受“云”的强大计算处理能力。在未来的几年里，云计算处理技术将渗透到人们生活的方方面面。



图1–5 云计算处理技术的构成

模仿人脑思维是未来大数据处理的重要方式。人工智能是实现计算机模拟人脑思维过程的技术基础。现阶段的人工智能主要通过计算机的算法模型实现模糊处理，这种处理方式体现出一定的智能化，但与人类智能的灵活性、目的性、创造性相距甚远。比如，AlphaGo（阿尔法围棋）就是利用深度学习来学习人类棋谱，通过蒙特卡罗树搜索模拟人脑来进行优势选择。模仿人脑思维的人工智能关键是汇集足够多的有效数据，再利用算法对这些数据进行处理。要实现模拟人脑思维的人工智能数据处理方式，必须借鉴人脑的先进结构和机制，这将是未来大数据处理技术的重要方式。

以社会学为原点进行多维度数据分析是未来大数据处理的思维模式。社会学是运用科学的方法研究社会与人类行为的综合性学科，它包括社会经济学、社会心理学、社会历史学、社会行为学，甚至包括法律、伦理、宗教等内容。社会学研究本身的综合性和研究对象的复杂性决定了其更多地强调相关性，这与大数据的方式相同。在大数据时代，人类的行为和活动可以通过数据化的方式进行研究和分析。其实不仅仅包括社会活动，任何人的私密活动也囊括在数据之中，人类的活动在被逐渐地数据化。造痕[注]指人类社会生活留下的痕迹，人类通常拿这些痕迹作为证据，广泛应用于考古学、历史学和社会学的研究中。数据化的证据就是一种数据。研究这种数据对于我们研究过去的社会结构、人类行为、文化等诸多问题都具有重要的意义，并能通过对以往人类社会和人类活动的分析，对未来社会发展趋势做出预测。某种意义上说，这种对社会的研究穿越了时空的界限，而这正是大数据应用的重要领域。

因此，以人的行为为原点，以数据社会学为基础，探索和研究大数据处理的规律，是未来大数据发展的重要思维方式。

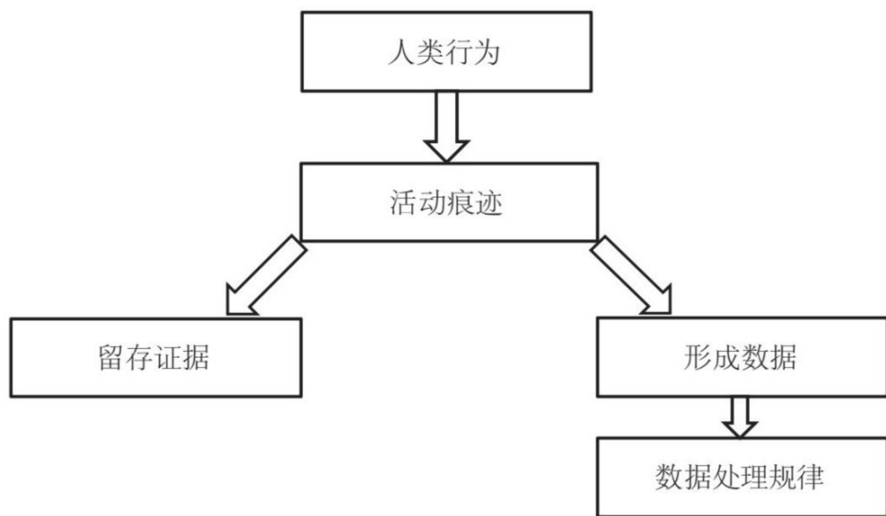


图1–6 人类造痕数据化处理过程

智能碰撞是未来大数据处理的重要方向。大数据发展的高级阶段就是数据能够进行自激活，不再需要人类发出指令，数据在流动过程中通过数据质点之间相互碰撞、相互影响、相互融合，实现智能碰撞和人机交互。这种计算机处理技术标志着数据进入最高级的阶段。在数据的智能碰撞过程中，每一个数据质点自动进行潜计算或者休眠计算，若达到激活条件，则会产生数据热点，同时也将自激活并进行热点数据处理和热点逻辑计算，并且持续重复此过程，将信息传递给其他数据点。各热点根据数据分析计算，显现出热点数据背后的规律，最终形成预测数据，这个过程就是数据自激活的过程。

（三）块数据解决方案

“条时代”的数据观。大数据强调“把数学算法运用到海量数据上来预测事情发生的可能性”，是“数据、技术、思维三足鼎立的产物”。[注]大数据观的核心是量化一切，认为世界的本质就是数据，万物皆可被数据化。数据从作为事物及其关系的表征走向了主体地位[注]，成为独立于人类社会的客观数据世界。借助智能终端、物联网、云存储、云计算等技术手段，人类获得海量数据，并从中选取对自身有价值的信息进行处理。目前人类对大数据的使用，更多的是对特定领域或行业内纵深数据的集合，这对于反映本领域或行业的规律具有很大的价值。但

是，这种条数据的处理方式将数据困在一个个孤立的链条上，相互之间不能链接起来。在思维模式上，条数据是传统人类研究范式的数据化体现，是对单独领域的深化，不同领域间彼此割裂、互不融通。类似于目前人类的科学研究被划分为众多学科，对科学的研究只是在特定研究领域内不断地深化和推进，而学科之间彼此没有联系。反映在价值观上，条数据是一种利我的、闭合的观念，这里的我，既是特定的主体，也是特定的领域，是虚拟化的“我”。这种条数据的数据观具有很多天然的缺陷。比如，根据分类，数据存储在不同部门，各个部门的数据仓库架构、数据技术也可能不同，政府、企业的数据互相无法打通，产生数据孤岛；大量的数据被政府部门、企业尤其是大型企业掌握造成数据垄断；对事物的预测仅限于利用特定领域的有限数据，造成数据预测失真，这也是产生海量信息悖论的一个重要原因。

“块时代”的数据观。打破数据孤岛和数据垄断需要新的数据观，这种新的数据观就是块数据。块数据是基于复杂科学提出来的新的研究思维范式，是一种数据化的整体思维，强调相关性而非因果性，用数据规律补充了单一的因果规律，实现了唯理论和经验论的数据化统一。块数据强调“更多”（全体优于部分）、“更杂”（杂多优于单一）、“更好”（相关优于因果）[注]，注重整体性、多样性、关联性、动态性、开放性和平等性，通过对复杂性科学思维的技术化处理，使得复杂性科学方法论变成了可以具体操作的工具，实现了定性定量的综合集成，使人文社会科学等曾经难以数据化的领域实现了定量研究，是一种全新的大数据方法论。在价值观上，块数据强调开放共享、跨界融合，是一种利他的、共享的观念。

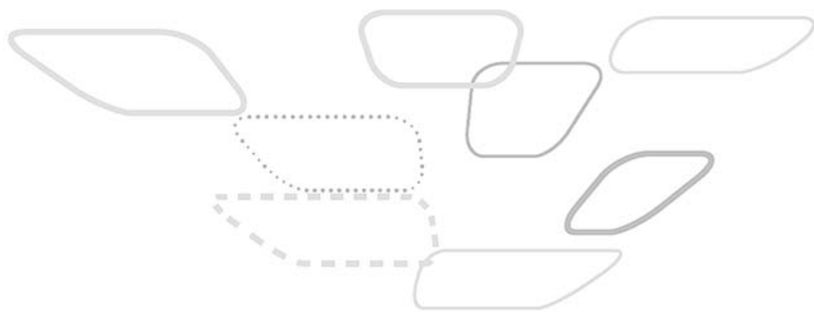
从大数据走向块数据是必然趋势。大数据强调开放共享，但在“条时代”，大数据的发展面临共享难度大、垄断程度高、融合能力差、应用价值低以及安全风险大等一系列制约因素。块数据是具有高度关联性的数据在特定平台上的持续集聚，既是数据集聚的结果，也是数据聚合的过程；既包括点数据、条数据，也包括面数据；既有数据空间的填充、空间数据的重构，也有聚合过程中的组构和组构过程中的聚合，同时还伴随新数据的汇集和原有数据组合后衍生数据的产生。这种聚合是持续进行的，并伴随着数据的不断更新。这个过程既是块数据自我重构和自我修复的过程，也是对条数据组合、纠偏、选择的过程。同时，块数据的聚合是具有高度关联性数据的聚集，数据的价值关联性决定了块数据的组构过程中并非简单堆砌，而是呈空间网状分布，并具有

明显的网线、节点、脉络及其自身内在的逻辑运行规律。块数据自身所具有的平台化、关联度高、集聚力强、价值密度高、开放性等特点，决定了块数据可以挖掘出数据更高、更多的价值，它将推动大数据发展进入条数据和块数据融合发展的新阶段。打破“条”的界限，让大数据实现在“块”上的“条”融合，是未来大数据发展的必然趋势。

---

1. 赵言舟，赵磊：知识是思维活动的基本构件，《政工学刊》，1998年第1期。
2. 王竹立：新建构主义——网络时代的学习理论，《远程教育杂志》，2011年第2期。
3. 香农：通信的数学理论，《贝尔实验室技术杂志》，1948年第27卷。
4. 马克·布尔金：《信息论》，王恒军、嵇利安、王宏勇，译。北京：知识产权出版社，2015年。
5. 张茉楠：大数据国家战略推动“数据驱动经济”，《南方都市报》，2015年11月6日。
6. 彭立志：《基于数据引力的分类方法及网络异常检测模型的研究》，济南大学硕士论文，2006年5月。
7. 复杂性科学主要包括：早期研究阶段的一般系统论、控制论、人工智能；后期研究阶段的耗散结构理论、协同学、超循环理论、突变论、混沌理论、分形理论和元胞自动机理论。
8. 麦肯锡环球研究院：《大数据：创新、竞争和生产力的下一个前沿》，2011年5月。
9. 维基百科，http: //en. wikipedia. org/wiki/Big-data。
10. 纳斯达克为世界各地组织的重要业务提供及时、相关和可信的数据，帮助其获得信息经济竞争优势。
11. 维克托·迈尔–舍恩伯格：《大数据时代》，周涛，等译。杭州：浙江人民出版社，2013年。
12. 李国杰：《对大数据的再认识》，北京：中国科学院计算技术研究所，2015年6月。
13. 肖锋：论信息技术时代的三大认识论悖论，《创新》，2016年第1期。
14. 樊月龙：大数据基本架构，http://blog.sina.com.cn/s/blog 933e5f350101 kuyk._html，2013年7月30日。
15. 高书国：大数据时代的数据困惑——教育研究的数据困境，《教育科学研究》，2015年第1期。
16. 艾尔·巴比：《社会学研究方法》，邱泽奇，译。北京：华夏出版社，2009年。
17. 维克托·迈尔–舍恩伯格：《大数据时代》，周涛，等译。杭州：浙江人民出版社，2013年。
18. 黄欣荣：大数据时代的哲学变革，《光明日报》，2014年12月3日。
19. 黄欣荣：大数据时代的哲学变革，《光明日报》，2014年12月3日。

第二章
结构化数据、抽象化数据与暗数据



一位美国记者在一篇文章中说，大数据是什么并不重要，重要的是大数据改变了人们对世界的看法。这种改变，以及改变所带来的变化，就是大数据的本质。那么这种改变是如何发生的，或者说正在或将要发生哪些变化，正是我们要探索和讨论的重大课题。

对于这个问题的研究，至关重要的是大数据的范围远远超出人类认知，大数据既包括通过人脑和电脑能够识别的结构化数据和非结构化数据，也包括只有人脑才能识别的抽象化数据，还包括至今人脑和电脑都无法识别的暗数据。这些数据彼此关联，叠加交错，相互融合，使我们对大数据的认识和研究变得更加复杂，对大数据处理技术带来了巨大挑战。正是在这一背景下，块数据作为大数据未来发展的一个解决方案，显得意义重大。

第一节 结构化数据

（一）数据的记录、测量和计算

人类进入信息时代后，数据内涵不断扩大，可以代指一切保存在数据库里的信息，其中包括大量的文本、图片、视频、音频等，也就是非结构化数据。这些非结构化数据本身就是信息，其来源不是对世界的测量，而是对世界的记录。

数据结构是数据存在的形式，也是数据的一种组织方式，其目的是为了提高算法的效率，也是达到数据语义化的最好方式。对数据结构的研究起源于程序设计。数据结构是研究数据元素之间抽象化的相互关系和这种关系在计算机中的存储表示（即所谓数据的逻辑结构和物理结构），同时要对这种结构定义相适应的运算，设计出相应的算法，而且确保经过运算后所形成的新结构类型仍然保持不变。数据结构化的过程，包括对数据的记录、测量和计算。随着计算机科学与技术的不断发展，计算机处理的数据也由纯粹的数值发展到字符、表格、图形、图像、声音等具有一定结构的数据。相应的，数据结构化也经历了无结构化、结构化、面向对象三个阶段。无结构化阶段的数据是纯粹的数值，数据之间的关系主要是数学公式或数学模型；结构化阶段的数据开始涉及并处理非数值；面向对象阶段，计算机开始对具有相同属性和方法的数据进行结构化处理。

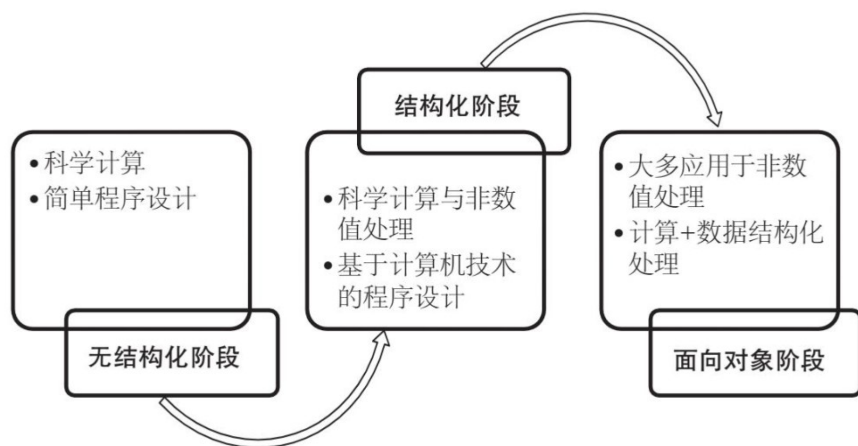


图2–1 数据结构化发展历程

结构化数据是可以被电脑识别的数据。结构化数据指具有一定结构性、可以划分为固定的基本组成要素、能够通过一个或多个二维表来表示的数据。结构化数据一般存储在关系数据库中，具有一定的逻辑结构，可以用关系型数据库的表或视图来表示，具有精确度高、统计方便、操作简单、可迅速建模分析等特点。结构化数据的出现极大地方便了人们的日常工作与生活。人们工作生活中涉及的数据信息存储在预先建立好的关系型数据库中，再把数据按业务分类，并设计相应的表，然后将对应的信息保存在相应的表中。使用关系型数据库来管理结构化数据是目前最好的一种方法，比如ERP（企业资源计划）、财务系统、HIS（医院管理信息系统）数据库等。

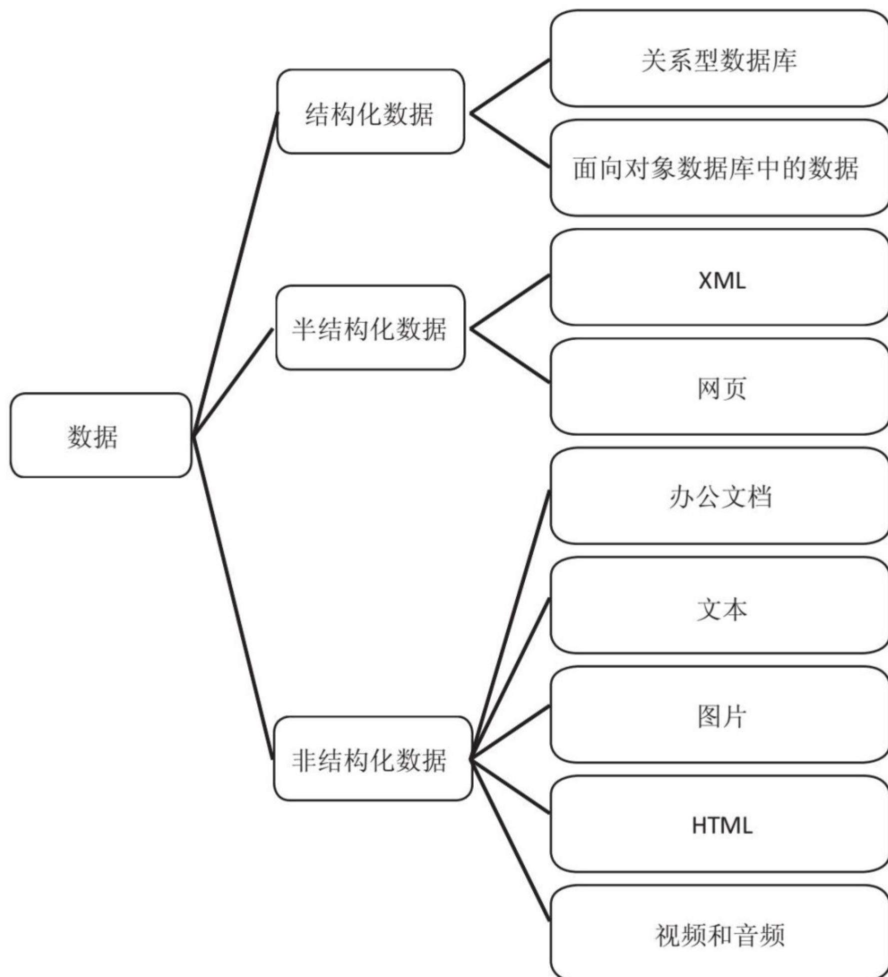


图2–2 数据存储结构分类

（二）结构化数据与非结构化数据

非结构化数据是结构化数据的反面。与结构化数据的形态不同，非结构化数据是指没有一个预定义的数据模型或不是以一种预先定义好的方式进行组织的数据。这种数据是直接按照学科方式分组分类，不能直接用数据库二维逻辑表来表现，在形态上主要包括我们日常的文本、图片、音频、视频、电视流等。

非结构化数据＞结构化数据。仅从计算机是否可直接识别的角度来看，目前非结构化数据的容量要比结构化数据大得多，数据产生的速度更快，来源更广泛。近年来，中国网民每年都会产生大量数据，尤其移动端的爆发式增长产生了大量非结构化数据。比如，人们在聊天时

的记录、邮件、发布的图片、语音以及视频过程中产生的数据通常都是非结构化数据。据国际数据中心调查显示，企业80%以上的数据都是非结构化的，且每年以60%的速度呈指数级增长。

从非结构化数据到结构化数据。非结构化数据不易被测量和计算，所以为了便于存储、分析和利用，有时需要将非结构化数据转化为结构化数据。特别是在互联网时代，由于计算机芯片、硬件的发展，计算机从主机型发展到普适计算阶段，移动智能设备、各种传感器、RFID（射频识别）标签、可穿戴设备被大量应用，越来越多的非结构化数据被计算机记录下来，但还不能处理。比如随着以脸谱网、推特网等为代表的社交媒体的出现，互联网成为人类“行为数据”集聚的平台，在微博、QQ（即时通信软件）、微信等交流软件中产生大量图片、视频、音频等非结构化数据。对这些非结构化数据价值进行分析和挖掘的过程，就是非结构化数据向结构化数据转化的过程。解决非结构化数据如何结构化的问题是大数据处理的关键，但是对非结构化数据的处理比结构化数据更为复杂和艰难。目前，从非结构化数据向结构化数据的转化大多采取“非结构化数据——半结构化数据——结构化数据”的方式。

注 比如，我们可以实现用计算机对连续文本进行语义理解，根据语义内容再将数据内容转化为结构化数据，利用成熟的关系数据库技术对其进行管理和分析预测。又比如，网站的商品信息数据就是非结构化的，商品颜色、规格、风格、包装等诸多信息都属于非结构化数据，我们要利用计算机将其一一识别并按照结构化数据的特点填放到“表格”中，才能使商品数据变得易于分析处理。

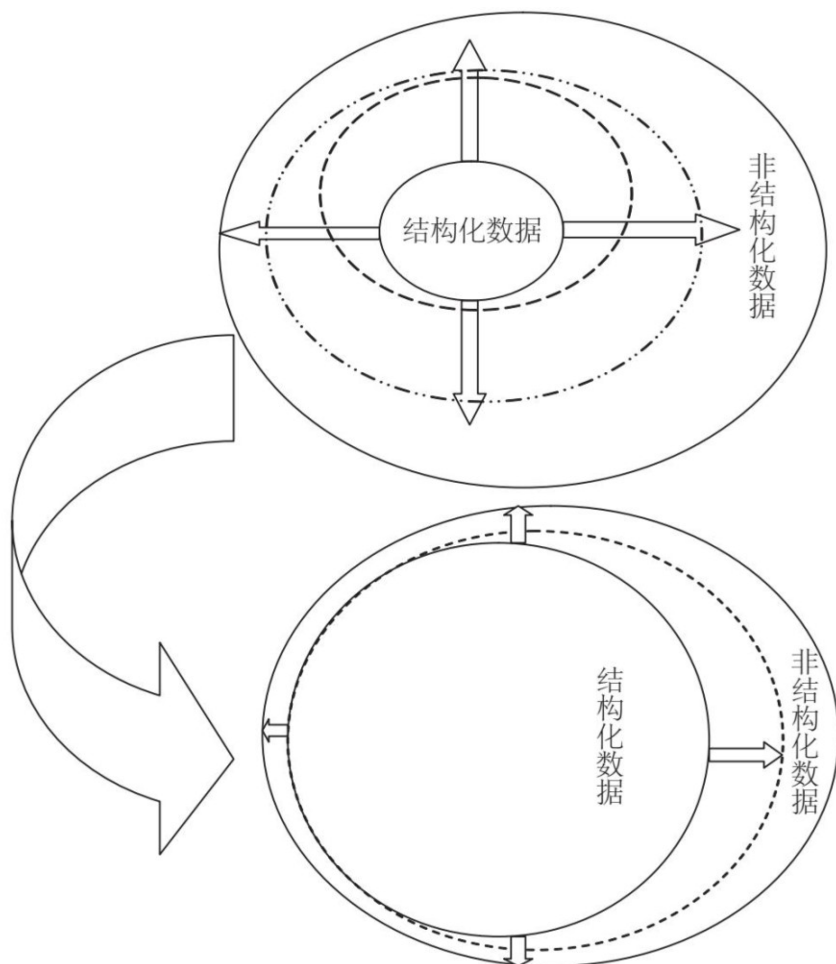


图2–3 非结构化数据向结构化数据的转化

（三）结构化数据的精确性与局限性

结构化数据更能反映事物的“本真”。数据作为一种客观存在，其准确性主要体现在数据能否准确反映客观事物本身。从计算机对结构化数据的分类过程来看，首先要对数据进行“预处理”，首要环节就是要对数据进行清洗。数据清洗是通过检查数据的一致性、处理无效值和缺失值等，对数据进行重新审查和校验，核心任务就是要清洗掉不符合规则要求的数据，主要包括残缺的数据、错误的数据、重复的数据三大类。通过数据清洗，尽可能地确保结构化数据的完整性、准确性，最大限度全面客观地反映事物本身。

结构化数据精确性虽高，但也存在局限性。一方面，结构化数据存

在一定的安全风险。大数据时代，大量非结构化数据仅用关系型数据库对其进行分析会耗时耗力。在非结构化数据结构化的过程中，需要借助云计算、分布式处理技术、存储技术和感知技术等新兴技术才能实现对大数据的采集、处理、存储到形成结果并进行使用。这些技术本身就隐藏了众多的安全风险与挑战，增加了结构化数据的风险性。因此，对数据安全漏洞的挖掘分析以及安全风险的综合管控就更加关键。另一方面，结构化数据并不能涵盖大数据时代的所有数据。目前，我们所研究的结构化数据和能够被结构化的非结构化数据，都是以电脑识别为基础的，但是在大数据时代，还存在大量不能被电脑所识别的数据，比如只能被人脑所感知但是还不能被电脑记录、测量、计算的抽象化数据和目前不能被人脑、电脑所感知或识别但是实际存在的暗数据。同时，即便是能够被电脑所记录的非结构化数据，由于目前的数据处理技术还不能有效地处理多维数据，数据在结构化过程中还面临着诸多的技术瓶颈，比如容量有限，可扩展性低，建设和运营成本高等，这在很大程度上也制约了结构化数据的发展和使用。

第二节 抽象化数据

（一）人脑感知

抽象化数据是介于结构化数据与暗数据之间的，人脑通过视觉、听觉、味觉、触觉、嗅觉等感官很容易感知到，但是不容易被电脑识别的数据。在现阶段，只有部分抽象化数据能够被电脑记录，但还不能实现测量和计算。人体的每一个器官都是外在世界信号的“接收器”，通过刺激，人体可以接受外界信号并将其转换成为感知信号，再经由自身的神经系统传输到人脑。伯克利大学生物解剖学家玛莉安·戴蒙在《人脑彩书》中将人脑称为“整个地球，或许整个银河系最复杂的物体”。与计算机相比，人脑的神经元相当于处理器。虽然速度不及电脑，但人脑数百亿个神经元就相当于一个极端庞大的分布式计算系统。神经元之间相互连接，连接处的突触就是人脑的存储器。这种类比电脑处理器与存储器的紧密相连的系统结构，就是人脑复杂的神经网络，它使人脑具有对外部世界信息强大的感知能力。人机工程学研究结果显示，人类所能感知到的信息，80%以上是由视觉器官提供的。目前计算机的虚拟感知系统建模研究大部分都是针对视觉系统来进行的。如防止互联网空间恶意注

册的验证码的应用，就是利用机器无法识别的一些人为扭曲、奇形怪状的字符原理来设计的，这些字符只有通过人脑才能够识别和处理，是抽象化数据的典型。

计算机是人脑感知的延伸。20世纪70年代，计算机科学的一个重要分支——人工智能出现，并被认为是世界三大尖端技术之一。人工智能是研究、开发、模拟和扩展人的智能的理论、方法、技术和应用的学科注，很重要的一个方面就是通过计算机模拟人的某些思维过程和智能行为，比如学习、推理、思考、规划等。未来世界一定是一个高度智能化的世界。注大数据打开了人类迈进人工智能时代的一扇大门。某种意义上，我们可以把人工智能看成“大脑”。在这个体系中，物联网是这个“大脑”的感觉神经系统，通过传感器感知，使人与人、人与物、物与物之间的交流变得可能，实现人类社会、信息空间和物理世界的高度融合。云计算是这个“大脑”的中枢神经系统，物联网的传感器和互联网的用户通过网络线路和计算机终端与云计算进行交互，向云计算提供数据，接受云计算提供的服务。大数据是这个“大脑”接收的外部信号，分散在整个系统中，体量巨大，借助云计算对这些信息进行处理，形成可以输入和输出的有价值的信息，它是智慧和意识产生的基础。互联网是运动神经系统，通过互联网，持续不断地向“大脑”提供大数据，供中枢神经系统进行决策。



图2–4 大数据、云计算、人工智能和物联网的关系

近年来，人工智能技术快速发展，并在某些方面显示出人脑不可比拟的优势。比如，“深蓝”和“沃森”计算机就在一些技能方面战胜了人。但它们依靠的并非自身逻辑思维能力，而是其内部计算的速度和存储空间。“沃森”计算机参加比赛时存储着4TB的文本数据库，覆盖了维基百科和名词网络（Wordnet）等互联网资料库。“深蓝”计算机在与冠军卡斯帕罗夫的象棋博弈中，依靠的是电脑每秒分析两亿种象棋路数的模型算法。注但是，目前计算机的运行效率并不高，在能耗上，“沃森”计算机拥有90台服务器，耗电8.5万瓦，而成人大脑的功耗只有20瓦左右。又如，谷歌人工智能AlphaGo与李世石的围棋大战，最终人工智能以4：1战胜人类。但这主要不是得益于逻辑推理的归纳与演绎，而是归于大数据的完备性和多维性。更为重要的是，人工智能虽然擅长数学计算，但当面对语音识别、人脸识别、情绪察觉等场景时却逊于人类。

（二）经验的抽象

经验的抽象，就是对人类发展历程中感知的信息以及在这一过程中进行的思考、判断、推理等人脑活动进行提炼和虚拟化。

对人类经验本身的抽象。经验是人们在同客观世界接触的过程中，通过感觉器官获得的关于客观世界的现象和表面联系的认识。亚里士多德认为，人类认识的对象是客观世界的具体事物，即实体，因此需要依靠感觉经验才能实现和完成这种认识。但是经验是停留在感性阶段的认识，尚未提炼和上升为规律或者理论。人类的日常经验判断是由视觉、听觉、味觉、嗅觉和触觉这5种感觉传递的信息所支配的。正如2004年，杰夫·霍金斯在《人工智能的未来》中提出的“一个算法”假说，认为人类诞生之初，大脑的神经网络还只是一张白纸，它必须通过每天不停地接受来自视觉、听觉等神经的海量信息，让神经网络从中自动总结出规律。对人类经验的抽象，就是要对大量的感性经验认识进行总结、分析、提炼出内在规律，并上升为理性知识。著名批判社会学家米尔斯在《社会学的想象力》中指出，抽象经验主义太注重研究程序和方法，而无法对实质性问题进行深入思考和研究，表现了高度的形式和空洞。虽然米尔斯从认识论的角度对抽象经验主义持质疑态度，但却提出了一个十分重要的观点，即对经验的抽象是一种重要的研究范式和思维方法。

对人脑活动经验的抽象。思考、判断、推理等人脑思维活动本身

是有经验的，也就是头脑中的思维模式。在现实生活中，人类对外界事物所做出的认知和反应是基于既有的认识和经验的。在认识新的事物时，可以根据已有的经验进行推理和归类，从而迅速做出判断，这个过程就是人脑思维的过程。对人脑活动经验的抽象，就是对思考、判断、推理等人脑思维活动经验的虚拟化。信息论认为人脑是信息处理器，信息处理过程就是信息的输入、编码、加工、存储、提取和使用的过程，在这个过程中伴随着人的知觉、记忆、思维和态度等。人工智能和计算机是通过对人脑活动过程的模拟，实现对人脑内部信息加工过程的逻辑分析，本质上就是对人类已有经验的提炼和虚拟。

经验抽象化的社会效应。经验抽象化产生的直接社会效应就是改变人传统的心理活动和行为模式。就如马克思所讲的，技术一旦为人所发明，如同存在一个某种自主性的“他者”对人自身产生影响，发挥着种种反主体性效应。在大数据时代，计算机通过对人类经验的抽象化处理，可以实现视觉、听觉、触觉等感官的模拟。这种模拟三维空间的虚拟世界，会让人暂时性地脱离现实世界，沉迷于电脑所创建的环境中。当人们参与到大数据的平台上时，受到平台吸附功能（召集力和广播效应）的影响，产生羊群效应、从众心理、蚁群效应，会形成群体的参与感、归属感、认同感，人们自身的经验常识和行为习惯会“暂时休眠”，形成不同于个体本身经验的行为心理和价值认同。

（三）抽象化数据的转化

通过对抽象化数据的加工处理，即编码、存储、提取、遗忘，实现计算机对数据的可识别、可感知，就是抽象化数据的转化过程。

抽象化数据的转化，目的是把人脑感知的数据转化为电脑可识别、易识别的数据。目前，对数据离散化处理是抽象化数据转化的重要方式。数据离散化处理是指计算机将人的感知活动合理地“分割”或“碎片化”，把对感知的整体性、连续性分析转化为“个体性”的离散化处理，将人的感知解构为无数计算机可识别、可记录的数据，再通过计算机数据分析系统，对这些数据进行重构，以达到对人脑感知活动的模拟。

数据建模是对抽象化数据进行处理和应用的重要方式。在这方面，最为典型的就是神经网络模型。神经网络模型是由大量、简单的神经元广泛地互相连接而形成的复杂网络系统，是一种基于模仿生物大脑的结构和功能而构成的数据处理系统。神经网络具有大规模并行、分布式存

储和处理、自组织、自适应和自学能力，特别适合处理需要同时考虑多种因素和条件、不精确和模糊的数据。[注]神经网络模型在模式识别、图像处理、智能控制、机器人等领域得到广泛的应用。比如，为了实现人机交互中文本和声音两种数据的自由转化，就需要建立包括自然语言处理、语音识别、声音合成等的“统计语言模型”，才能使计算机听懂人类语言，开口说话。

第三节 暗数据

（一）从暗物质到暗数据

暗物质是一种比电子和光子还要小的物质，不带电荷，不与电子发生干扰，能够穿越电磁波和引力场，是宇宙的重要组成部分。目前科学界对暗物质的认识达成了一些共识：暗物质来自宇宙大爆炸；暗物质无法用任何光学或电磁观测设备直接观测到，宇宙中95%以上是暗物质和暗能量，暗物质占26.8%；暗物质密度小、速度快，难以捕捉；除了引力之外，暗物质和它自己以及其他物质不发生任何作用，是促使宇宙膨胀时在自身引力下形成特定结构的首要物质类型。

世界的本源是物质还是数据？毕达哥拉斯学派认为数是宇宙万物的本源，提出“万物皆数”，“数是万物的本质”，数是“存在由之构成的原则”等观点。在他们看来，数为宇宙提供了一个概念模型，数量和形状决定一切自然物体的形式。数有量的多寡，还有具体的几何形状，认为自然界的一切现象和规律都由数来决定，必须服从“数的和谐”。由此，世界万物都被数据化了。中国古代的《易经》数相理论也认为，宇宙起源于数，宇宙的本质是数，宇宙之“数”由天文数（0到9）和地理数（一到十）构成。宇宙间万事万物的变化其实是“数”的变化。[注]我们认为，“世界的本源是物质”与“世界的本源是数据”内在统一，数据本身就是一种客观存在，同时又是对现实世界的记录，是对现实存在的一种数据映射。现实宇宙中一切的存在都会以数据的形式进行记录，但受人类认知能力和现阶段科学技术的影响，人类对于数据的认知就如同对宇宙的认知一样，还存在大量的不能认识或者尚未认识的范畴，我们称之为暗数据。

暗数据不仅仅是对暗物质的类映射。暗物质是客观存在的，就一定

会以数据的形式进行映射和记录。但是和物质与数据一一对应的映射关系不同，暗数据对暗物质的数据映射是一种类映射。对于暗物质的数据记录，超出目前人类所能认知的范畴，因此，这类暗数据还属于不能认知的暗数据。同时，暗数据又不限于对暗物质的数据映射，尽管暗物质的量十分庞大。从人类社会发展的阶段性来看，目前还存在大量可以被人脑或者电脑认知，但是尚未发现或记录的数据，这部分数据也是暗数据。

“黑暗”是相对的。如果把不能识别或者尚未识别的数据称为“绝对”暗数据，还有一种“相对”暗数据，这类数据是本身已经被人脑或者电脑识别并存储在各种相应的系统中，但是由于没有与其他数据发生价值关联（孤立数据），或者价值关联没有被呈现，造成未被使用、分析甚至访问。这类数据本身不是黑暗的，而是尚未被“点亮”。“相对”暗数据是我们研究暗数据的一类重要数据，也是目前我们研究和利用暗数据的重要类型。

在整个“数据宇宙”中，人类现阶段真正能够发现、认识和使用的数据的量只是很少的一部分。正是由于暗物质和暗数据的存在，世界变得难以预测，以不可确定性对抗确定性成为未来发展的一种趋势。

(二)暗数据的能量

暗数据也有引力。与明数据一样，暗数据也是由数据质点构成，在数据质点的作用下产生数据引力，并且暗数据的数据引力在显性和隐性状态中的能量级是一样的。暗数据通过数据引力波在不同数据引力场里的聚合、裂变，对明数据、暗数据以及所在的引力场都会发挥作用，释放能量。与明数据显性的数据引力不同，暗数据的数据引力是隐性的力，只有通过自激活或他激活，暗数据的价值才能显现出来。在暗数据被激活之前，这种引力波对事物的作用和影响不大，但会形成一定的干扰。

发现和挖掘暗数据使大数据对事物的分析预测更精准。大数据的核心在于对数据的分析、挖掘和预测。目前，人类运用大数据对客观事物进行预测的能力是有限的，影响预测能力的关键就是暗数据。我们对于事物的数据进行的预测分析，都是以明数据为基础的，大量客观存在并切实发挥作用的暗数据尚没有进入大数据分析的范畴，使得大数据的预测能力受到限制。如同索罗斯所说数据的“反身性”，参与者的思维与进

入的情景相互联系与影响，彼此无法独立，认知与参与处在永恒变化状态中。[注]这种变化就是由大量暗数据导致的。发现暗数据、挖掘暗数据会使大数据对事物的分析预测更加精准。《易经》最初就是用来预测的，《易经》分析的对象是数相，本质上也是数据，但强调的是“全息数据”[注]，其范畴可以理解为明数据和暗数据之和。《易经》对事物的预测是从暗物质、暗能量层面，采用特殊的预测方法和流程进行占筮分析和预测，包括起卦、析卦、断卦，实际上就是对信息的获取分析和形成结论的过程。通过起卦，获得特定事物的卦符号，再分析卦符号中所蕴藏的数相信息，进而对事物加以预判。[注]《易经》预测的过程综合考虑了事物的显性数据和隐性数据。我们在研究大数据的过程中，应该注重对暗数据的研究和分析，扩充明数据信息量，提升数据质量，减小预测值的偏离度，进而使相关的分析、判断和预测更趋准确。

（三）暗数据的识别

暗数据是大数据范畴中，除了明数据之外的数据。暗数据来自大数据的“阴暗面”，但是更代表了大数据中光明的一面。[注]与暗物质不同，暗数据一旦被识别和使用，就不再是暗数据。也就是说，暗数据是可以转化为明数据的，并且这种转化是单向的。暗数据一旦变成明数据，就会被记录和保存，而不会再变为暗数据。某种意义上可以说，数据的发展史就是暗数据不断转化为明数据的过程。

暗数据与明数据不是割裂的对立面，而是具有某种关联性的。运用明数据的关联性来识别暗数据是目前最为有效的识别暗数据的方法。不是所有的暗数据在特定的时空范围都需要被识别，只有与特定的明数据相关联时，暗数据的识别才有意义。当然，这种关联性是间接的或者是隐性的。简言之，就是要利用数据的多维性，来分析数据背后的数据或者是与之相关的数据，进而实现对暗数据的识别。这就类似在刑侦破案过程中，找不到犯罪嫌疑人或者不能确定其具体的身份（暗数据），只能借助已有的信息（明数据）锁定相关的联系人（数据的数据），通过对他们的了解分析、推理判断（关联性识别），进而锁定或者发现犯罪嫌疑人（识别暗数据）。当然，对暗数据的识别要比这个过程复杂、抽象得多。用于识别暗数据的关联性数据，可能与明数据、暗数据存在于不同的数据引力场之中，也可能是需要多重间接关联性才能获得的数据。总之，尽管暗数据对大数据具有重要的作用，但是人类对暗数据的识别还有很长的路要走。

（四）暗数据的激活

英国学者维克托在《大数据时代》中指出："数据的真实价值就像漂浮于海平面的冰山，人们看到的只是冰山一角，绝大部分却藏于海洋的表面之下。"无论是"绝对"暗数据还是"相对"暗数据，其共同的特点就是内在价值尚未被发现并发挥作用。就如同英特尔中国研究院首席工程师吴甘沙在"大数据的黑暗之海"论断中指出的，数据之于数据社会，就如同水之于城市或者血液之于身体。暗数据本身就是"水"或者"血液"的重要组成部分，只有让暗数据"活"起来，才能发挥暗数据强大的能量，人类才能真正迎来"大数据的黄金时代"。

暗数据的激活是一个打破数据孤岛、发现暗数据引力即价值关联性的过程。研究激活的暗数据对其所在的数据引力场发生的作用，有助于提高通过数据发现潜在规律、预测未来的能力。

暗数据的激活包括自激活和他激活两种。暗数据自激活的过程类似于人类中枢神经系统运行过程。借助数据引力，暗数据与其他数据具有某种价值关联，在数据引力波的作用下，构成一个复杂的自组织、自适应、自流程的数据系统，也就是数据引力场。在数据引力场中，暗数据与其他数据特别是核心数据间的价值关联不断累积，在达到特定的临界点时，暗数据的价值就会显现出来，实现从暗数据到明数据的转化，也就是暗数据的自激活。暗数据的他激活是暗数据借助外部力量，主要是通过数据搜索、数据挖掘等技术，让暗数据进入人脑或计算机的认知范围，并进行数据记录和处理的过程。

（五）暗数据的剔除

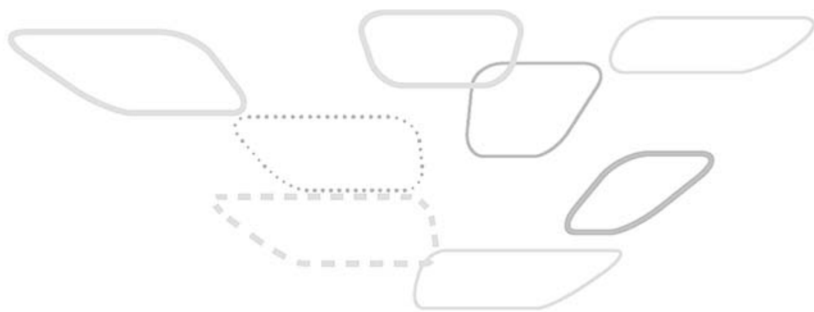
暗数据的剔除不等于删除。数据删除是指利用现有的技术手段，将数据库中已用的数据人为或者自动删掉，这一过程中，数据的删除是永久的、不可恢复的。数据剔除是在对数据进行分析处理的过程中，将部分数据排除出去，不纳入分析的范围。与数据删除不同，数据剔除本身不改变数据库中的数据容量。被剔除的数据依然客观存在，在必要的条件下可以重新使用。暗数据的剔除是对激活过程中或者处于激活状态的暗数据的选取过程，其目的是为了提高数据分析的客观性和准确性。暗数据的剔除就类似于人脑自动筛选数据的过程。人脑本身就是一个庞大的数据库，内含着大量的暗数据，在接受外界刺激过程中，人脑会自动选取对刺激有效的暗数据，对于无效的数据自动屏蔽。

暗数据的剔除对象包括无效数据和干扰数据。无效数据主要是指不在特定数据处理范畴的数据，其存在与否不影响数据分析的结果。暗数据的数据量十分庞大，但不是所有的暗数据都会对特定的数据分析具有必要的价值关联性，没有必要的价值关联性的数据就是无效数据，这部分数据的存在与否不会对分析结果产生任何影响，但是会增加数据处理的负担。干扰数据主要是暗数据在激活过程中产生的扭曲数据，包括误差数据、重复数据、异常数据等，这些数据的存在会对数据的处理造成干扰，影响数据分析和结果预测的准确性。对于暗数据的剔除技术，可以参考计算机在对结构化数据预处理过程中采用的数据清洗技术。

---

1. 陈琳琳、李建林：《数据结构与算法》（C语言版），北京：清华大学出版社，2015年。
2. 万里鹏：《非结构化到结构化数据转换的研究与实现》，西南交通大学，2013年。
3. 王汉华、刘兴亮、张小平：《智能爆炸》，北京：机械工业出版社，2015年。
4. 杰夫·霍金斯、桑德拉·布拉克斯莉：《智能时代》，李蓝、刘知远，译。北京：中国华侨出版社，2014年。
5. 黄俊杰：像人脑一样感知，《第一财经周刊》，2013年9月18日。
6. 胡瑞敏、徐正金：人工神经元网络的智能神经元模型，《电子学报》，1996年第4期。
7. 黎斌：《〈易经〉“数相”与“大数据”》，中国大数据产业观察网，2016年3月18日。
8. 这里的“绝对”本身是一个相对概念，是从静止的角度来理解的。随着人类发展和技术进步，不能识别或者尚未识别的数据也有可能变成明数据。
9. 索罗斯：《金融炼金术》，孙忠、侯纯，译。海口：海南出版社，1999年。
10. 黎斌：《〈易经〉“数相”与“大数据”》，中国大数据产业观察网，2016年3月18日。
11. 黎斌：《〈易经〉“数相”与“大数据”》，中国大数据产业观察网，2016年3月18日。
12. 佚名：《数据的阴暗面：什么是暗数据？为什么暗数据很重要？》，36大数据网，2015年10月22日。

第三章
块数据的本质



对块数据的认识和理解，是随着对数据奥秘的探索和对大数据价值的发现而不断深入的。从某种意义上说，块数据和大数据是相伴而生的。将大数据作为参照，区分大数据和块数据，是我们遵循的一个基本理论研究方法。块数据的概念提出以来，对“块数据是什么”的描述就从来没有停歇过。在此基础上，我们期望对块数据的概念进行更加确切的把握和更加精炼的说明。基于对科学技术、人类思维模式和社会组织方式的前瞻，我们认为“块数据是高度关联的各类数据在特定平台上的持续聚合”。就像数学上对未知数的设定赋值，为块数据赋予一定的意义或形象，将有利于块数据更广泛地交流和应用。

我们并不想止步于概念阐释，而是力图形成一个开展块数据研究和运用块数据思想的坐标系、参照系及对其规律的探寻。基于此，我们将开展数据社会学范式的思考，形成对块数据发展的共同认知。

第一节 块数据的定义

（一）块数据集聚

大数据的价值在于海量和关联。注到目前为止，大数据的应用都是以条数据呈现的。实际上，条数据已经实现了数据的集聚，即在一定条件下同类型、同领域数据的集中，这种集中可称为数据的指向性集

聚。从传统企业存储的会员卡数据、传统金融业存储的银行卡数据，到互联网企业存储的电子商务数据、互联网金融存储的行业数据，再到各政府部门信息化管理存储的卫生、教育、交通等民生数据……其集聚和使用的数据指向都被限定在某个行业或领域，都是一种数据指向性集聚。这种集聚实现了同类数据的关联，使得人们能够清晰地掌握某个领域的整体状况和最新动态，进而预判的精准性提高，生产生活成本降低……数据的使用提升到一个新的层次。但与此同时，指向性集聚也带来了数据单一、数据封闭、数据垄断等问题，突破数据指向性集聚的瓶颈是块数据发展的价值所在，也是块数据形成的基础。

不管是条数据还是块数据，数据集聚都是它们共同的运行规律。如前所述，由于数据引力波的存在，数据集聚会形成一定规律：当数据的指向性集聚达到一定规模的时候，数据引力场就会形成。这个数据引力场超越时空，促进数据实现由低能量向高能量的转化，并产生一种巨大的数据引力，在这种数据引力的作用下，数据的关联性集聚就会产生，数据从“条”到“块”的转化将得以实现。

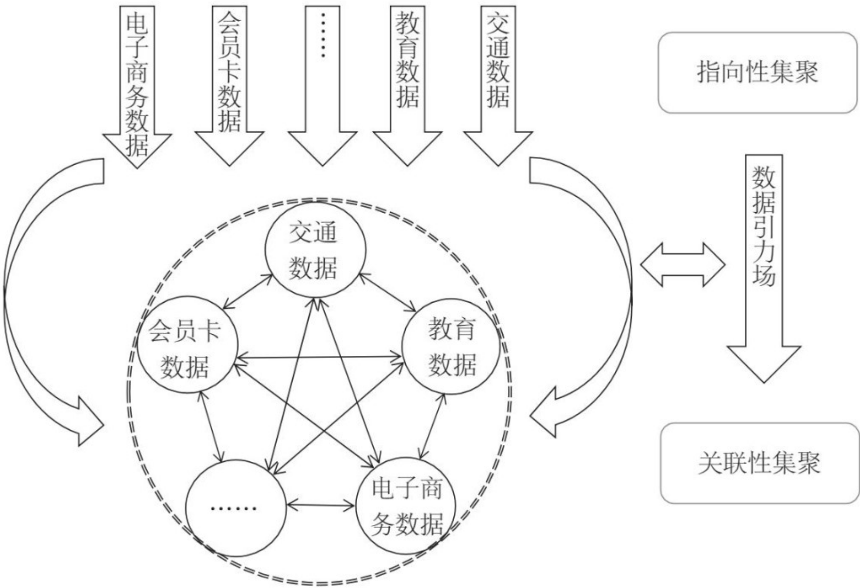


图3–1 指向性集聚与关联性集聚

与条数据不同，块数据是具有高度关联性的各类数据在特定平台上的持续聚合。这种持续聚合的实质是一种关联性集聚。关联性集聚与指向性集聚的区别在哪里？首先，指向性集聚是同一领域、同一类型数据的集聚，关联性集聚是各类数据的集聚。其次，指向性集聚实现的是数

据的垂直关联，关联性集聚实现的是数据多维的跨界关联，是一种内在的、紧密的高度关联。这种高度关联不是数据的简单相加，而是数据的聚合。数据的关联程度越高，聚合的能力就越强，持续更新的速度就越快，基于此形成的关联分析和关联挖掘的深度、广度将不断拓展。再次，关联性集聚是在指向性集聚基础上实现的持续集聚，如果说指向性集聚带来的是规模效应的话，那么关联性集聚将实现激活效应，并循环往复，实现质变。



图3–2 关联性集聚与指向性集聚的区别

例如，目前贵阳市一些先进社区通过“社会和云”平台建设，搭建了社区居民与居委会交流反馈的社交平台，该平台不仅采集了社区居民、家庭、楼宇的基本数据，而且融合了民政、公安、教育、金融、生活、养老等社区领域内的数据，同时还通过实时监控和反馈采集了动态数据。接下来，各类数据的聚合将带来价值再造，形成一个共享、开放的社区块数据。

块数据的关联性集聚并不局限于某个物理空间或行政区域，而是在特定的平台上实现的，这个平台既包括特定的物理空间，也包括虚拟空间、操作环境，比如软件等，还包括某种工具，比如阅读器等。块数据的关联性集聚，将打破传统的信息不对称和物理区域、行业领域对信息流动的限制，通过对不同类型、不同领域数据的跨界集聚，极大地改变信息的生产、传播、加工和组织方式，进而给各个行业的创新发展带来新的驱动力，推动各个领域的彻底变革和再造。注

（二）块数据关联

关联关系是块数据的特质，也是客观的存在。我们面临的难题是如何把关联关系数据化，并进行分析与挖掘，从而实现精准预测。关联关系最典型的说法是大家熟知的蝴蝶效应。它是由爱德华·诺顿·洛伦茨于1972年发表的论文——《蝴蝶效应》中提出的，并由此产生了可与相对论、量子论相提并论的混沌学。

混沌理论是一种兼具量化分析与质性思考的方法，用以探讨动态系统中无法用单一的数据关系，而必须用整体、连续的数据关系才能加以解释和预测的行为，比如人口移动、化学反应、社会行为等。[注]

混沌系统具有其鲜明的特性。比如随机性，它表现为关键性因素不确定、难以把握；再比如敏感性，混沌系统的运动轨迹是从离散到连续、从低维到高维的，敏感性表现为在这个运动过程中对外部刺激会产生强烈响应。混沌系统还具有分维性，这种分维性会带来时空无穷缠绕、折叠和扭结，并构成具有无穷层次的自相似、自组织结构，只要数值计算的精度或实验的分辨率足够高，就可以从中发现内在的关联性和有序性。

混沌理论认为，“一切事物的原始状态，都是一堆看似毫不关联的碎片，但是这种混沌状态结束后，这些无机的碎片会有机地汇集成一个整体”。[注]混沌状态是否结束取决于外部刺激何时出现，或者说引爆点何时出现。如果能够提前预测这个引爆点，就能够实现对事物的控制和对事件的应对。

块数据就是数据世界混沌理论的代表，它是一种将量化分析与质性思考高度结合的分析方法，其任务是透过无规律、不可预测的数据寻求事物的内在规律，实现精准预测。块数据是一个典型的混沌系统，它从表象上看混乱无序，但实际上富有深层次的规律性，经过长期及完整分析之后，可以从中得出某种规则。它的特点是对外部信息的刺激十分敏感，但内在具备某种（或几种）可被引爆的点。就像蝴蝶扇动翅膀的运动，导致其身边的空气系统发生变化，并产生微弱的气流，而微弱气流的产生又会引起四周空气或其他系统产生相应的变化，由此引发一个连锁反应，最终导致其他系统的极大变化。[注]蝴蝶扇动的翅膀就是整个连锁反应的引爆点。又如，金融大鳄索罗斯的一个小小举动就掀起了涉及整个东南亚的金融风暴，这个引爆点表面上是索罗斯的举动，但本质

上是金融体制漏洞中的某个“数据”。关于时空旅行的例子也能说明这一点，如果在时间之河的上游投入一枚小小的石子，到下游的时候就会变成一场巨大的洪水灾难，发生改变的就在于石子这个引爆点。块数据的运行规则是在离散化数据集聚的基础上，实现高度关联，数据会出现有机聚合，并形成整体关联的数据关系，当某个引爆点发生的时候，块数据能够迅速捕捉并准确预测事物发展的趋势与规律。

数据集聚是探究关联关系的前提。数据过于分散，不便于进行关联比较，当数据实现持续集聚并达到海量级时，关联关系的分析与关联规则的挖掘才能实现，并能够对数据的关联度进行衡量。关联度表征的是数据之间的关联程度。外在关联往往是低度关联，内在关联则表现为高度关联。在量子力学中，有一个关于互补性的概念，它是指在宏观世界不会同时出现的两类现象，如波动性和粒子性，在描述微观现象、解释实验时却是互相补充、不可缺少的。块数据的关联在一定意义上就发挥着这样的作用。

关联关系的分析是第一步，是对各类数据之间的外在关联或显性关联的比较。什么是各类数据？它可以从不同的角度进行划分。按人类社会形态来分，未来的数据可以分为个人数据、法人数据、社会数据、国家数据；按数据来源来分，各类数据包括涉及人、事、物的数据，也包括人以外各类组织的数据；按数据时态来分，既包括过去的统计数据、现在的实时数据，也包括未来的预测数据；按数据表现形式来分，既包括结构化数据，也包括非结构化数据，还包括抽象化数据和暗数据等。对各类数据的识别、采集、存储、传输、使用，是进行块数据关联关系分析的基础。

块数据并不止步于关联关系的分析，它更加强调关联规则的挖掘。相互关联的数据隐含着内在的关系和模型[注]，关联规则的挖掘将发现一些潜在的、有价值的、高度关联的内在信息，并找到引爆点。关联关系的分析是基于单维、单层数据的处理，比如用户购买的物品；关联规则的挖掘则是基于多维、多层数据的处理。换句话说，关联关系的分析是处理单个属性中的一些关系；关联规则的挖掘则是处理各类属性之间的某些关系，揭示“由于某些事件的发生而引起另外一些事件发生”的规则。关联关系的分析和关联规则的挖掘，共同完成了块数据这个混沌系统运行的一个完整周期。

在大数据研究中，有一个被广泛传颂效仿的“啤酒与尿布”的案例。

美国沃尔玛超市基于其庞大的数据仓库系统，得出了一个有趣的发现，啤酒与尿布摆放在一起可以促进二者销量的双双增长。结合大量的消费行为调查和分析，一个隐藏在“啤酒与尿布”背后的美国人的消费行为模式开始显现，并被应用于营销策略的制定之中。这是一个典型的数据关联关系分析的例子。随着块数据的产生，“啤酒与尿布”的故事还可以继续。一方面，可以基于美国人的消费行为，对美国的家庭组织模式和社会运行规则做进一步的分析，通过数据关联规则的挖掘，使分析更能接近本质和规律；另一方面，块数据的应用将比超市购买问题更加广泛和复杂，对于引爆点的预测，将使得对数据关联规则的应用领域扩展到管理决策等更大范围。

我们更加熟悉的是淘宝网的关联营销，它通过关联关系的分析将事物、产品、品牌等所要营销的要素进行关联，从而实现对客户需求和广告推送的精准匹配，实现深层次的多面引导。目前，这种关联关系的分析仅止步于淘宝自身的应用，如果基于块数据的关联规则的挖掘能够实现，数据将成为淘宝最具价值的资源，也许那时候的淘宝将更为强大。

（三）块数据价值

块数据通过特定平台集聚高度关联的数据，进而对数据进行关联关系的分析和关联规则的挖掘，以实现块数据价值发现、重构和创造。从大数据到块数据的演变或转化，就是从具有一定关联到高度关联的过渡，高度关联的数据比普通意义上的大数据的价值密度更高。

大数据是以容量大、类型多、存取速度快、应用价值高为主要特征的数据集合，但大数据的高价值应用是基于某个领域、某个行业的内部数据的集聚与关联。按照迈克尔·波特的产业价值链理论，价值最大化不仅取决于企业的内部价值链，更取决于在一个更大的价值系统中，一个企业的价值链同供应商、销售商以及顾客价值链之间的连接，也就是跨界关联。[注]相较大数据而言，块数据的价值就是为产业发展、公共服务、社会治理提供了一个更大的价值系统。换句话说，块数据的最大价值在于应用，是解决现实世界中的各种难题的关键。从商业、社会到政府，都可以基于块数据找出解决方案，这正是块数据的价值所在。

大数据推动新业态发展，大家熟知的Zipcar（一家网上租车公司）、Airbnb（一个社区型租房平台）、Uber（优步）、滴滴、神州等企业的出现正是这种新业态的最好例证，它们形成的是以协同消费模式

为主导的共享经济价值链。而基于更大的价值系统，块数据将催生新的全产业链。也就是说，基于块数据的发展，各产业部门之间依据内在的关联规则，形成新的关联关系，会将我们带入众联、众包、众创、众筹的平台化模式，实现资源再配置，从而对传统产业结构和产业体系进行颠覆性解构和重构，形成新的全产业链。

维克托·迈尔–舍恩伯格在《大数据时代》中指出：“大数据时代的经济学、政治学、社会学和许多科学门类都会发生甚至是本质上的变化和发展，进而影响人类的价值体系、知识体系和生活方式。”基于更大的价值系统，在标准化和技术化的基础上，块数据将催生新的全服务链，人们可以选择更加个性化的生活方式，获得更加精准化的服务。

大数据正逐步成为现代社会基础设施的一部分，并成为政府实施管理的重要手段，就像公路、铁路、港口、水电一样不可或缺。例如，谷歌曾经根据搜索记录的庞大数据资源预测出了流感的传播趋势，大数据在管理当中的应用给政府带来了震惊与反思。基于更大的价值系统，块数据的多层、多维的关联关系分析和关联规则挖掘，将比谷歌基于单层、单维的搜索记录分析来得更加精准。这种精准预测的实现，将进一步提升政府治理能力、创新政府服务模式，催生新的全范围治理链。块数据量化和预判的特性对社会生活的影响将是全方位、多领域的，特别是对激发各个社会组织、群体、个人参与社会公益，凝聚社会共识具有不可替代的作用，可以说，其在社会治理中的应用价值与经济领域的应用价值同样巨大。注

第二节 块数据的基本属性

（一）主体性

主体性是一个哲学问题。在人类历史的发展过程中，人与物的关系一直是人们探寻的一个核心问题。古希腊哲学家毕达哥拉斯认为“数是万物的本源”，用数据来观察世界，不可避免地也需要回答人与物的关系问题。如果说条数据是围绕着“物”而产生的（这里的物包括企业的产品和服务），那么，块数据则是因“人或组织”而存在的，它呈现出主体性。大数据用人的思维观察和解释数据，而块数据则是用数据思维去观察和解释人的行为规律。块数据的框架将推动人类行为与数据的交互影响以

及人类自身的进步。

块数据以人为原点，围绕着人去记录静态数据、行为数据、意识数据，这些数据反映了在块数据中人作为主体性的存在。

人的属性与静态数据。静态数据包括人的姓名、年龄、性别、职业等。这种静态数据主要记录的是人的基本属性，不仅是人实现自我认识的必要条件，也是外部对人形成认识的基础，但静态数据难以区分人与物的本质差别。

人的能动与行为数据。行为数据是围绕着人的活动所形成的活动空间范围、活动行为轨迹、能动关系等数据。与蜜蜂筑巢、蜘蛛织网不同的是，人的行为具有社会性和能动性。人之所以成为主体，从根本上来说，就是这种具有能动性的创造活动决定的。人的行为数据反映了人与外部社会的关系，比静态数据更具价值。

人的超越与意识数据。意识数据是包括人的思想、意识、意愿等在内的数据。在自为、自觉的基础上，不断实现自我超越，是人主体性的最高表现形式。人的创造性活动是开放的，是面向未来的，当意识可以用数据的形式保存的时候，智能碰撞、人机交互将成为可能，人将实现一种新的对技术的超越和自我超越。

（二）高度关联性

高度关联性是块数据的本质属性。它主要表现为三个方面：首先，众多高度关联的数据在特定平台上集聚；其次，数据与数据之间构建了一种彼此连接、相互交错，集灵活性与网络性于一体的组织连接格局；再次，集聚于平台上的数据之间形成一种相互影响的关联机制。

块数据集聚的是一种高关联数据。这种数据有两个典型特征——影响力大和感应度强。当影响力大的数据发生变化时，就会通过数据之间的相互关联产生波及效应，在数据持续集聚中发挥更大作用。感应度强的数据在持续集聚中能迅速地感知其他数据的变化，并灵敏地应对与调节。高关联数据是数据解构与重构中的关键数据。

块数据形成了一种高关联格局。块数据的本质和规律隐藏在各种数据的相互关联之中。通过块数据认识问题、分析问题、解决问题，需要进行“关联”，构建人的关联、物的关联、人与人的关联、人与物的关

联、时空的关联，形成彼此连接、相互交错，网络状数据连接格局。

块数据运行的是一种高关联机制。块数据集聚与块数据价值实现，在理论上存在着相互影响的关联机制。数据高度关联为持续集聚的发展提供了必要的条件，并推动持续集聚向更高层次发展。数据的关联度越高，信息量越大，价值也就越大。

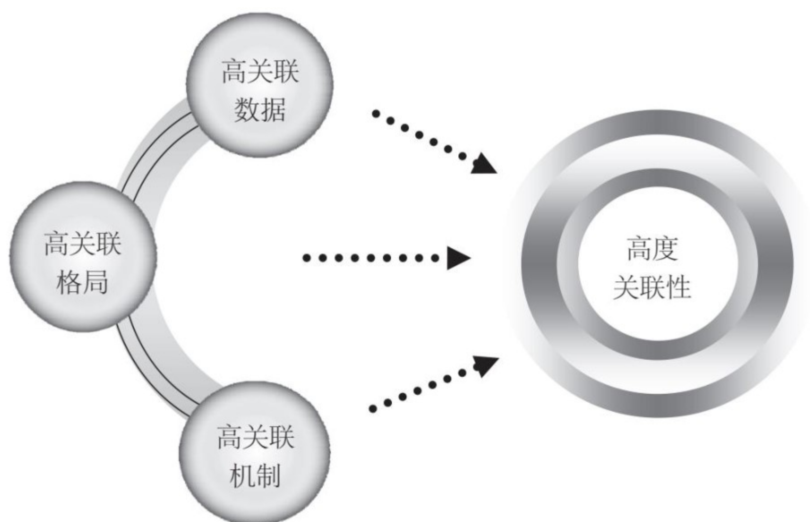


图3–3 块数据的高度关联性

（三）多维性

块数据是一个引爆点。相对于大数据，块数据呈现出立体多维性。这种多维性带给我们的是一种整体、关联、系统、灵敏地审视与适应世界的方式，使我们的视野不仅仅局限于孤立的单元或领域。块数据的多维性表现在4个方面。

无边界。块数据的多维性首先表现为无边界。块数据跨界融合了政府部门、商贸物流、金融保险、工业企业、社交网络、影视娱乐、设备传感器等各个方面的数据，形成了一个更大的数据有机体。在这个有机体中，各类数据本身是有边界的，但块数据的内部结构没有进行预先的边界设定与限制。这就使得块数据像动物细胞与动物器官组织的关系一样，尽管数据与数据之间存在“隔膜”，但并不妨碍数据之间畅通无阻地流动，数据能量之间快速便捷地传递，以此保持块数据的灵敏性与适应性。

不确定性。人们通常用“黑天鹅”来指代不确定、不可预测的重大事件。块数据就像是“黑天鹅”，其多维性带来了不确定性。块数据既包含结构化和非结构化的数据，还包括抽象化数据和暗数据等，抽象化数据和暗数据的分布范围以及状态是不确知、不固定的。因此，块数据的不确定性更多的是由抽象化数据和暗数据带来的。这种不确定的特质在意料之外，却又举足轻重，既是块数据的价值所在，同时也带来了更高的风险。

非线性。非线性是一个数学概念，用来描述变量之间的数学关系。非线性现象往往是复杂的、多维的、不确定的、千变万化的，非线性更接近客观事物的本质。从某种意义上来说，条数据是线性的，条与条之间具有独立性，互不相关；块数据是非线性的，不是一种简单的线性叠加。随着数据维度的不断提高，数据处理的难度越来越大。就像激光一样，当外加电压达到一定值时，受激原子将得到统一指令，发射出相位和方向一致的单色光，当数据持续集聚到一定程度，外部作用将使得块数据内部的各类数据相互作用，从而获得更加接近真相的规律。

超时空。块数据是智慧的数据，它必然要进入一个更高维度的时空。按照“数是万物的本源”这一基本认识，看不见的暗物质与暗能量也会附着对应的暗数据。由此带来的数据能量或质量将通过超越时空的持续聚合，产生时空扭曲或时空折叠，并形成数据引力。数据引力释放出引力波，高度关联的各类数据将实现自激活和他激活。在此基础上，通过智能碰撞和人机交互，块数据将使得预测未来成为可能。

（四）强活性

强活性是一种能迅速自主反应或促进其他相关反应的激活属性。块数据的强活性主要表现为两种状态，它们可相互作用形成连贯机制。

一种状态是自激活效应。快速更新的数据是最具价值的。块数据强活性的表现之一就是数据以更快的速度和频率进行自激活。这种自我激活是在高度关联的各类数据接收到外部信号之后的一种反应。外部信号就像钥匙，各类数据就像门锁，当外部信号和各类数据的运行规则高度契合的时候，数据就会实现自激活，数据重新连接、重新生成，以获得源源不断的数据应用价值。可以用一个形象的例子来说明外部信息如何发生作用。例如，当一株柳树上出现毛毛虫啃食树叶时，这柳树就会

分泌某种化学物质，给60米半径范围内的柳树发出信息。邻近的柳树收到这种报警信息后将分泌一种有毒物质，驱走毛毛虫。

另一种状态是他激活效应。块数据还具有一种类似于化学反应中的催化功能，在获得外部信号之后，其中某类强活性的数据会扮演催化剂的角色，能够提高或降低其他数据相互反应的速率，实现重新连接和重新生成，而其本身没有发生变化。

所谓的连贯机制是激活效应之间的相互作用。有一个形象的比喻是，一个是快速移动、时刻变化的河水，一个是稳定的、缓慢变化的河床，随着时间的推移，水流会慢慢地塑造河床，河床也会改变水流的方向。数据的自激活效应就像河水，数据的他激活效应就像河床，相互影响、相互作用，推动各类数据的共同演化。

（五）开放性

开放性是为了探寻事物内外的各种联系以发现规律。块数据的开放性是指数据与数据之间、数据与外部环境之间发生交互关系的属性。就像中子与质子相互开放，交换介子，形成原子核；原子核与核外电子相互开放，交换光子，形成原子；原子与原子相互开放，交换电子，形成分子；分子与分子相互开放，交换物质与能量，才能形成各种宏观物体，块数据的开放性是其实现集聚、关联并最后重构价值的基础条件。块数据的开放性表现为以下特征。

可介入。块数据的多维性和无边界决定了其具有可介入特质。高度关联的各类数据可以在没有任何障碍、没有任何限制的条件下自由流动，从而实现持续集聚。

可拓展。高度关联的各类数据在实现持续集聚的过程中，将在一个更大的预留空间内进行相互作用，实现自激活和他激活，且这种相互作用将不断循环往复，达到一种自发性、自组织和自流程状态。块数据的这种可拓展性，就像生物的开放性突出地表现为同化作用与异化作用两个方面。生物的同化作用，就是从外部环境吸收所需物质，转化为自身的构成物；生物的异化作用，就是分解自身的物质。

可输出。这是块数据存在的必要条件。在自激活和他激活完成之后，块数据的开放性将实现数据由内部到外部的传递，将为接下来的智

能碰撞、人机交互提供条件。

第三节 块数据与大数据的区别

（一）大数据的4V特征

在我们努力定义块数据概念和描述块数据属性的时候，依然需要回到大数据。如前所述，块数据和大数据是相伴而生的，把大数据和块数据加以区别，会让我们更加全面地认识块数据。

目前学术界将大数据的特征归纳为4V：数据容量大、数据类型繁多、商业价值高、处理速度快。

数据容量大：数据体量巨大，超出了计算机的处理能力甚至超出了存储能力。目前，数据存储容量单位已经从TB级别，升级到PB级别、EB级别，甚至更高。

数据类型繁多：数据来源复杂，数据结构复杂。以往的数据处理往往忽略细节处的非结构化数据，只抽取便于存储的以文本为主的结构化数据。随着越来越多的非结构化数据大量产生，非结构化数据占到互联网数据总量的80%，给整个社会都提出了挑战，对数据存储和处理能力也提出了更高的要求。

商业价值高：蕴藏巨大的价值，同时价值密度与数据量成反比。稀薄的价值分布需要海量数据支撑，同时需要强大的算法进行价值挖掘。以谷歌和百度为例，虽然作为两家公司的核心技术的搜索引擎，具有信息采集的功能，但是海量的检索结果中，到底哪些是有用的，是一个极其耗费精力的再选择过程。

处理速度快：高速采集数据、迅速给出处理结果。这也是大数据区别于传统数据挖掘最显著的特征。现代科学技术的进步使得数据处理能力已有了几何级数的提升，同时使得数据产生速度也呈现几何级数增长，二者是对立统一的。

（二）块数据的多维变量

比较大数据和块数据，二者最显著的区别是从大数据的“4V”发展到了块数据的“5V”，即多维变量（Variable）的出现，而这也是块数据集聚、块数据关联、块数据价值得以产生的原因和基础所在。块数据的多维变量引发的是数据使用分析从静止和孤立转变到运动和联系的趋势。正如“数学中的转折点是笛卡儿的变量”一样，大数据时代的转折点在于块数据的多维变量的出现。

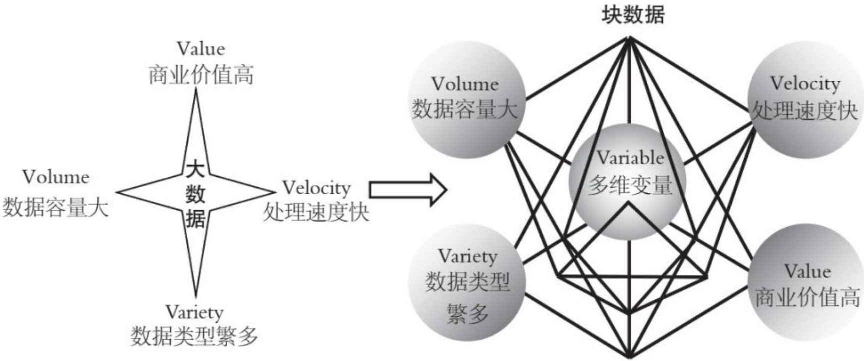


图3–4 块数据与大数据的区别

变量在数学中是指可以改变的数。多维是无限，块数据变量在空间和时间上没有限制。块数据的多维变量，将使得辩证法进入数据的分析与使用，形成块数据的辩证思维。变量会带来更大的不确定性、不可预知性和更高风险。但同时，这也使得块数据能够揭示各种因素的交互影响以及事物发展的内在规律。

常量和变量，表征的是不变与变。常量往往是静态的，变量通常是动态的。常量是相对的，有条件的；变量是绝对的，无条件的。常量和变量这对矛盾，是可以相互转化的。例如，在一定条件下，气温可看作常量。但是，冬天和夏天的温度不同，白天和黑夜的温度不同，北方和南方的温度不同，人与人感知的温度不同，这些情况下温度又是一个变量，它是随着时空和外部条件的变化而变化的。这就是所谓的“常”中必定有“变”，“变”中必定有“常”。恩格斯在《自然辩证法》中说：“整个自然界……都处于永恒的产生和消灭中，处于不断的流动中，处于无休止的运动和变化中”，绝对不变的常量是不存在的。相较于大数据，块数据是用更加辩证的思维，在高度关联的数据中获取变量。

德国物理学家沃纳·海森堡于1927年提出不确定性原理，该原理简单来说，就是在精确测定一个量子的位置或速度过程中，需要用到波长或

长或短的波，但对量子会产生扰动。越用波长较短的波精确测定它的位置时，速度测量就会越不精确；越用波长较长的波精确测量它的速度时，位置测量就会越不精确。这种不精确是由或长或短的波形成的扰动导致的，所谓的变量就是这种扰动因素。块数据致力于分析、挖掘影响事物发展基本趋势的变量。在许多变量中找出一些更基本的，但又无法直接感知的隐性变量，把握影响事物发展的扰动因素，使不可预知变成可预测、可预警、可预案。

有限和无限，反映的是数据在时间和空间上的辩证性质，而不仅仅代表大数据带来的巨大的数据体量。世界本质上是无限的，比如，太阳系是一个有限事物，是世界无限整体的一部分，在它之外的无限空间内存在无限的物质，但有限的太阳系和无限的物质之间是存在关联的，仅仅用太阳系的数据难以全面刻画无限的世界。在有限的时空中去认识事物更多的是基于因果关系，块数据的多维变量恰恰是为了找到太阳系和无限物质之间的关联关系。如果用量子纠缠理论的万物皆有默契来类比块数据，在数据皆有关联的基础上，块数据可以从多维时空聚合各类数据，并从中认取、分离出数据的潜在关联，从而全面认识和把握事物发展的规律与本质。

从常量到变量，从有限到无限，体现的是块数据的辩证思维。或许，用这种辩证思维也可以回答常常困扰我们的问题：“块数据是否是有形的块？”

第四节 块数据的范式转移

（一）范式与范式转移

在《科学革命的结构》中，托马斯·库恩首次提出了“范式”的概念。他认为范式是各种科学理论，以及与理论密切相关的哲学信仰、价值标准和研究方法。在开展块数据的研究中，库恩所用的范式研究方法给我们带来了极大的启示。

所谓“范式转移”，是指新的学术成果打破原有的假设或者法则。例如，当古希腊数学家、哲学家毕达哥拉斯提出地球是圆的，这种新范式就代替了旧范式，为人们的思想和行动开创了新的可能性。随着科学技

术的发展，特别是大数据时代的到来，自然科学与社会科学的融合发展成为必然。我们很难用某种传统的、独立的学科理论对块数据进行诠释，块数据社会学范式就是在这种背景下形成的一个新的学术成果，并将可能带来新的范式转移。

块数据社会学范式不是空穴来风，更不是空中楼阁。实际上，人类在社会学和数据分析中已经经历了一个漫长的进化历程，可以概括为4个阶段。第一阶段是社会集体经验阶段，人类从原始社会阶段摆脱自然条件的束缚，形成了人类的语言和信息共享，这是最原始的数据处理，实现了从无数据到有数据的飞跃。第二阶段是当人类文明出现以后，人类通过文明规律对社会的发展进行分析和评价，进而进行预测。比如，中国古代以史为鉴，通过分析历史的数据，预测未来发展的规律。这个阶段的数据囊括了历史、文化、生活、科学等方方面面。第三阶段是社会统计学阶段，这是在有了工业文明之后，人类为更多地了解人类与大自然的活动规律，人类与大自然发生了数据交换，产生了社会统计学，如天气预报、社会抽样调查，用样本数据做分析预测，这时，数据已经从科学数据发展到人文数据。第四阶段，人类将进入社会学、系统生物科学、云计算和人工智能融合发展的数据分析阶段。随着人类处理数据能力的提升，对数据的存储和计算将实现几何倍数的增长，而这种增长通常是无效的。由此，数据社会学分析方法将产生，以社会学为原点，融合计算机、云计算等技术工具进行多维化的数据分析，将带来新的社会范式和科学范式的出现。这种范式转移又必将引发新一轮数据革命，并导致新技术的产生和人们生产、生活、生存方式的巨大变革。

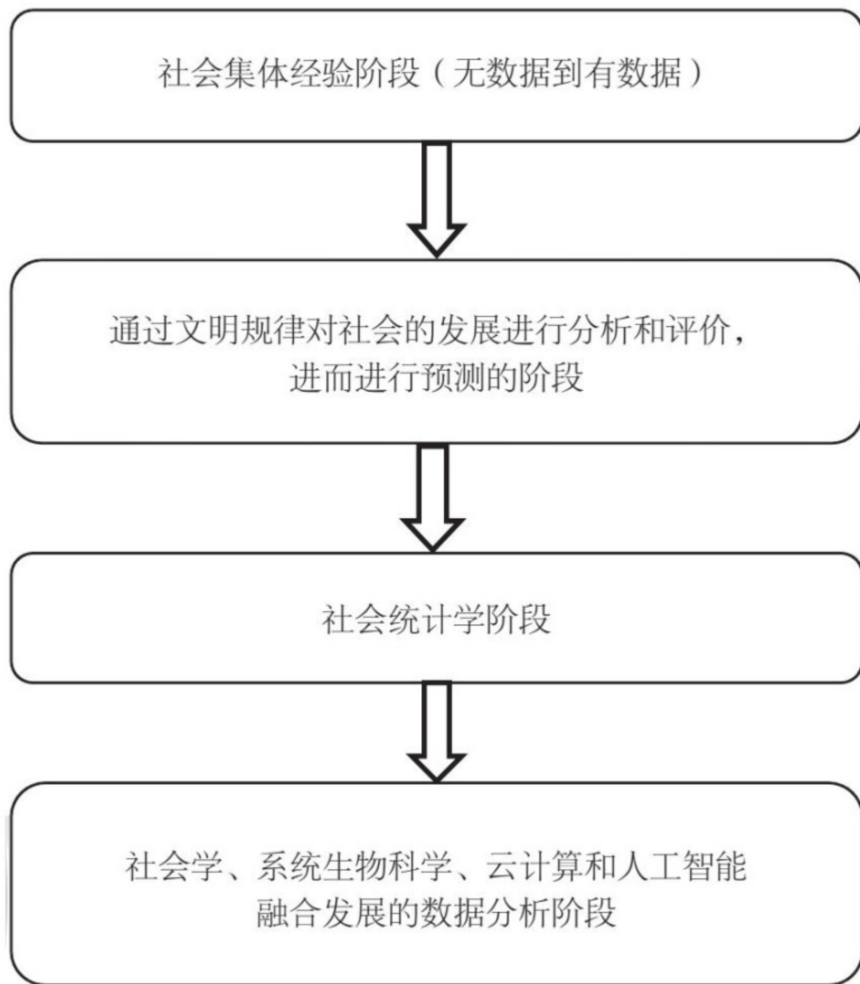


图3–5 人类社会学和数据分析的4 个阶段

（二）以人为原点的数据社会学范式

数据社会学范式具有复杂性这一特征，这种复杂性主要表现为整体性、工具性、多元性、互补性。

整体性表现为数据社会学范式强调用整体论和还原论相结合的方法，分析生命系统、社会系统、数据系统等复杂系统，它将带来传统科学方法的革新。数据社会学范式的整体性还体现为，它不仅仅是在理论层面上的探讨，而且从哲学信念、价值标准、研究方法等方面进行整体性思考。

工具性表现为数据社会学范式的出现，它克服了大数据以物为原点

的片面性，实现对人类行为与社会运行规律的客观认知与科学解释。任何新范式的提出，都是为解决某一方面或者某一层次的问题而产生的，数据社会学范式也不例外。

多元性表现为数据社会学范式是由多个学科共同支撑形成的。正如在社会科学领域中也未形成一个像牛顿力学或达尔文进化论那样占统治地位的统一范式一样，大数据的发展还未形成成熟统一的理论，不同的角度所透视的只是大数据发展的不同侧面，研究对象的多维性决定了数据社会学范式必然呈现出多元性和互补性。

我们试图从5个方面展开对数据社会学范式的研究，以此来描述我们对这一范式的基本模式、基本结构与基本功能的认识。

本体论。随着大数据的兴起，“世界的一切关系、一切活动皆可用数据表征”的观念逐步为人们所接受。世界的本质是数据，数据世界相对于客观世界而独立存在，数据世界的原点是物还是人，是数据社会学范式要解决的一个重要问题。块数据以人为原点，提出数据人假设，重构人与物的关系，实现人与人、物与物、人与物的全方位高度关联，实现人类社会的整体数据化。

认识论。数据社会学范式把辩证法运用于认识论，为解决大数据时代人类所面临的问题提供了一种具有普遍意义的认知和思维方式。块数据是通过本质与现象、内容与形式、原因与结果、必然性与偶然性、可能性与现实性等范畴，以数据化的手段揭示事物的本质联系，从而更加准确地把握客观世界的基本规律。

方法论。块数据强调世界的复杂多样性，更强调事物的潜在关联性，以此建立一种基于复杂理论的系统分析方法，并依据定量指标检测事物发展的引爆点，确定应对的方式方法。在这个分析系统中，智能体的行为以及人机交互将发挥至关重要的作用。块数据构建起的这种分析方法，在各学科、各领域可以进行普遍应用。

价值论。价值思维和价值取向的差异不仅影响人的思想和行为，也是决定一种科学范式将走向何方的重要因素。数据社会学范式不仅回答块数据的价值是什么，而且深入探讨应该怎样实现价值重构，试图按照这种范式推动块数据在商用、民用、政用等各个领域的应用。

伦理论。随着科学技术的发展和人工智能的不断升级，传统意义上的人与机器的关系将被颠覆，技术离人越来越近，甚至融入人体。机器开始出现可自演进的状态，甚至成为人类之外的另一种智能生命。人与机器在伦理层面上存在着文化交叠和意识碰撞，但问题远不像亚里士多德预想的那样简单，“无须手的指引，梭子能照原样织布，琴拨能照原样奏出旋律，一流的工人们也就不会想要仆人了”，人与机器不是简单的奴役与被奴役、控制与被控制的关系，或者正好相反。数据社会学范式尝试引导一种科技、伦理和人类自身相统一、相平衡的观念。

（三）块数据范式是一场新的科学革命

托马斯·库恩在《科学革命的结构》一书中指出，科学革命的结构是“科学——范式——反常——新范式——新科学”。《第四范式：数据密集型科学发现》将人类科学的发展定义为4个“范式”，每一个范式的出现都带来了一场新的科学革命。

第一范式是“实验科学”范式，这种范式主要是以记录和描述自然现象为主，如人类最伟大的发明应用之一就是钻木取火。第二范式是“理论科学”范式，这种范式主要是利用模型进行归纳总结，最具代表性的是牛顿三大定律、相对论等。第三范式是“仿真模拟”范式，这种范式主要是通过计算科学进行模拟仿真，推演复杂现象，典型案例是天气预报等。第四范式是“数据密集”范式，随着大数据时代的到来和人工智能的发展，在模拟仿真技术的基础上，通过密集的数据运算，人工智能将替代科学家从事更多工作，进行分析总结和规律抽离，就像人工智能AlphaGo赢了韩国围棋大师李世石那样。

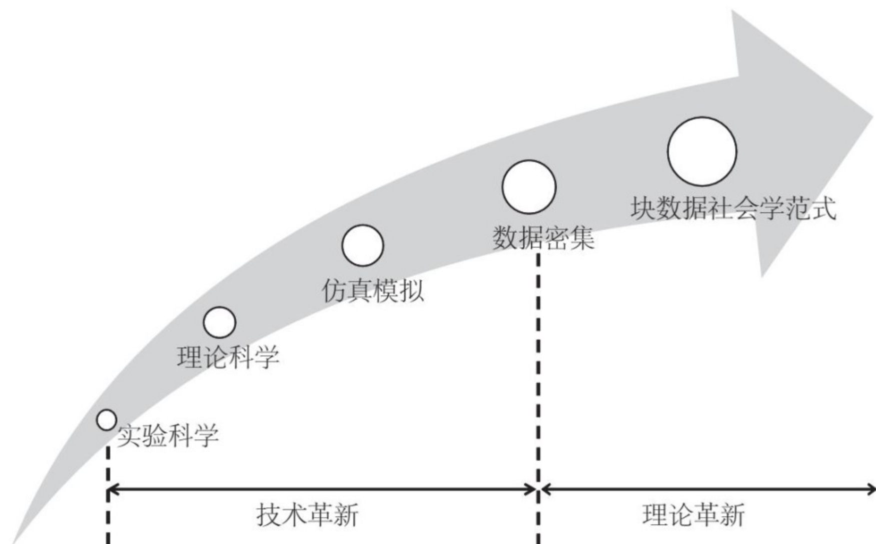


图3–6 科学范式与范式转移

爱因斯坦对科学的理解是："科学是人类寻求能把观察事实联系在一起的、最简单一致的思想体系的努力。"科学影响历史有两条途径：一条是技术革新，另一条是观念革新。当观念具有条理性的时候，就会形成思想总汇，并不断地影响人类发展。一定意义上说，块数据社会学范式的提出就是在技术革新的基础上形成的理论革新。相较于"数据密集"范式，块数据社会学范式的核心是人机交互、智能碰撞。这是一场意义深远而又科幻的科学革命，这场革命将改变我们的思维方式、改变世界上物质与意识的构成、改变我们的世界观，这场革命是顺势而为。

注

1. 大数据战略重点实验室：《块数据：大数据时代真正到来的标志》，北京：中信出版社，2015年。

2. 陈刚：块数据的理论创新与实践探索，《中国科技论坛》，2015年第4期。

3. 尹启天、马立丽：教学过程当中的混沌性，《今日科苑》，2008年第16期。

4. 吉志伟：混沌理论在建筑设计中的运用，《中外建筑》，2011年第6期。

5. 业茂：蝴蝶扇动了翅膀，《科学24小时》，2005年第2期。

6. 弗兰克·J·奥尔霍斯特：《大数据分析：点"数"成金》，王伟军、刘恺、杨光，译。北京：人民邮电出版社，2013年。

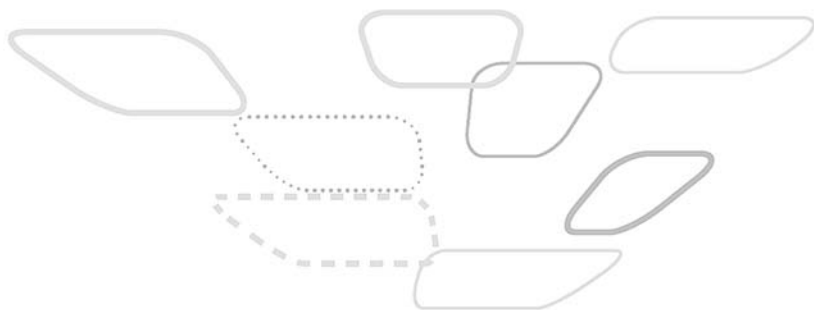
7. 汪斌：基于经营战略形态和产业价值链的经营战略矩阵设计探讨，《区域经济评论》，2012年第9期。

8. 大数据战略重点实验室：《块数据：大数据时代真正到来的标志》，北京：中信出版社，2015年。

9. 欧文·拉兹洛：《自我实现的宇宙》，杨福斌，译。杭州：浙江人民出版社，2015

年。

第四章
块数据模型



仅仅依据块数据的定义来描述数据的集聚、关联与价值，还不足以探究清楚块数据的本质，对块数据的认识也容易陷入单一性、概念化的局限。要科学系统地认识块数据，不仅要把握其概念与属性，更要从整体结构、内在机理和运行流程等方面构建一个科学模型。块数据模型的建立是块数据研究取得突破性进展的重要标志。

块数据模型将数据流动、聚集、关联、价值发现与再造的流程对应到平台化、关联度与聚合力三个环节。该模型指出平台化是数据流动与聚集的过程，关联度是建立数据连接与数据价值发现的过程，聚合力是价值再造的过程，在这三个过程中体现了数据的流动性、关联性、价值发现与再造的循环性。块数据模型的构建，将以量化的方式回答为什么块数据是大数据的解决方案，更进一步解释为什么块数据是大数据时代真正到来的标志。

当然，块数据模型的构建还处在初步阶段，所设计的测度指标未必能反映其全貌，有关平台化率、关联度系数及聚合力指数还缺乏具体的指标支撑。此外，对块数据的运行机制如数据迁移、数据群聚与数据引力的模式创新还不够完善，但这恰恰是推动块数据理论构建进一步走向科学化的重要基础。

第一节 模型理论与块数据模型

（一）模型理论

模型理论的两个核心概念是模型和建模。马格努斯·赫斯特尼斯在其模型理论中提到，模型是对真实事物的概念表征，是对象的替代物。[注]以色列学者斯勒及美国学者斯米特强调，模型不是对系统的真实描写，而是为了解释客观实在的某些方面而做出的一套假设；模型只是对客观实在进行形象化及做出解释的临时性变通工具。[注]MBA智库百科中提到，模型是指对于某个实际问题或客观事物、规律进行抽象后的一种形式化表达方式。关于建模，哈伦提出5个步骤，分别是模型选择、模型建构、模型证实、模型分析和模型拓展。[注]而现在模型构建一般采取模型准备、模型假设、模型构成、模型求解、模型分析、模型检验和模型应用7个步骤。模型最主要的特点是对客观事物、客观规律的抽象，最后回归于实际应用之中。

模型的表征方式并非单一的，其表征方式有数学表达式、流程框图或步骤、结构图、图表等，采取何种方式往往根据建模目标和模型类别进行选择。在模型的类别中，数学模型通常是一组数学表达公式，主要用于反映客观事物运行规律和变化发展的趋势；概念模型的表征形式有表达概念的示意图等；结构模型主要是反映系统的结构特点和因果关系，其中图模型[注]可以描述自然界与人类社会中事物之间的关系，是研究各种系统特别是复杂系统的有效方法。除此之外，模型还有程序模型、逻辑模型、方法模型、分析模型、管理模型、数据模型、系统模型等类别。

构建模型的意义在于描述复杂系统，阐述事物间的联系及增进人们对事物规律的认识。首先，模型构建有利于从细节上描述复杂系统，事物及系统本身具有复杂性，使用模型可以从细节上解析复杂系统，增强对相关细节的把控程度。其次，模型构建有利于揭示客观事物之间存在的联系。例如波特五力分析模型，揭示了供应商的议价能力、购买者的议价能力、潜在竞争者进入的能力、替代品的替代能力、行业内竞争者现在的竞争能力之间的组合变化与行业利润潜力变化之间的关系。最后，模型构建有利于人们更好地透过现象探究本质，更为系统地认识和理解客观事物之间存在的联系及其产生的影响。

（二）数据焦虑与块数据模型

焦虑的产生往往在于人们无法预测与掌控未知事物，同时也在于人

们无法理解与应用已知事物。大数据时代，人们的数据焦虑主要是对“见不到”与“见到得不到”的焦虑，“见不到”的焦虑表现为对未知领域探索的焦虑；“见到得不到”的焦虑主要表现为如何从得到的海量数据中快速获取有效数据及如何对海量数据进行处理分析，从而发现数据关联，再造数据价值。

未知带来焦虑。D·H·劳伦斯认为“未知”有三个特征：第一，未知是无限的，并超越了普通的世界；第二，未知崇高、神秘且神圣；第三，未知给予生命意义，它是通往不朽的道路。[注]千百年来，人类对未知领域的探索从未间断过，并在很多领域取得了突破性进展。但是宇宙的奥秘是无穷的，人们未知的领域很大，即便是下一秒，人们也不可预知会发生什么，不可控的因素太多，这种不可预测性与不可控性会造成集体焦虑。

近似无限的可获取数据量和有限的数据处理能力之间的矛盾，带来了“见到得不到”的焦虑。第一，快速获取有效数据的焦虑由两个方面的原因引起，一方面是每时每刻都会产生体量巨大的数据，这会造成数据超载现象，让人们疲于在海量的数据中快速获取有效数据；另一方面是信息热点更新极为快速，人们的专注和思考能力被撕成碎片。正如克里斯蒂娜·罗森所说：“当我试图在电脑上长时间阅读的时候，我的眼睛难得片刻休息，视线总是跳来跳去，分散精力的东西太多了。”[注]信息热点容易淹没在海量的数据中，这会影响人们快速获取有效数据。第二，大数据时代的主要矛盾是无限增长的数据体量和有限的数据处理能力之间的矛盾。大数据时代最关键的是数据处理能力。大数据不等于大价值，我们需要通过数据处理等手段获取数据价值，然而，目前的数据处理能力远远跟不上数据产生的速度，人们由此产生数据焦虑。

块数据模型构建是解决数据焦虑的有益探索。构建块数据模型是为了更加精准地理解和把握块数据的内涵和本质，要解决的问题主要有促进数据流动、建立数据连接、发现数据价值、再造数据价值，其中再造数据价值是需要解决的核心问题。

（三）块数据模型要素分析

块数据是高度关联的各类数据在特定平台上的持续集聚。它包括多层含义，首先，块数据是具有内在价值关联的数据，按照人为设定和自适应的规则持续集聚；其次，块数据是实现价值发现和价值再造的过

程；最后，块数据通过持续聚合得到具有较高价值密度的数据集合、价值集合与应用集合等多维度组合结果。我们把块数据理论的具体实现称为块数据模型，用公式表示为：

$$K=S(V,P)$$

其中，K代表块数据，S代表平台化，V代表关联度，P代表聚合力。

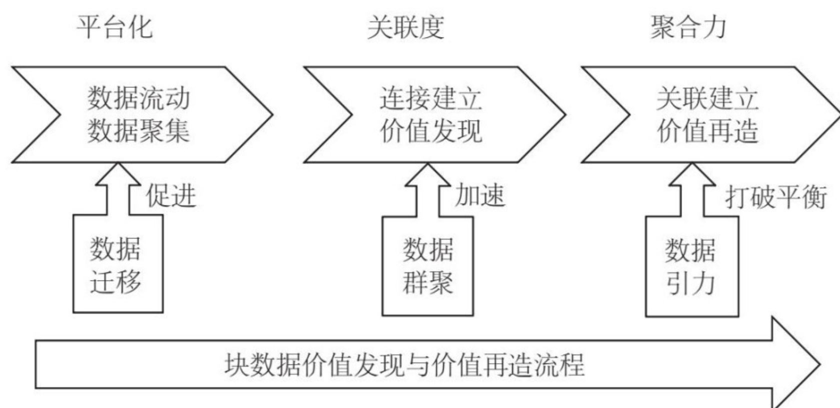


图4–1 块数据价值发现与价值再造流程

平台化、关联度与聚合力是块数据形成的三个重要环节。第一，平台化是各类数据通过流动实现聚集的动态过程，是衡量数据是否充分聚集的一个重要标志。平台化可以用平台化率进行测度，主要刻画数据集聚的程度，是数据是否完成平台化的判别机制。第二，关联度是在平台化的基础上，建立数据连接与发现数据价值的动态过程。关联度强调的是数据的跨界与融合，数据通过不断地与其他数据建立连接，从而发现价值。关联度可以用关联系数进行测度，主要刻画数据之间的相关程度。第三，块数据具有强活性，它打破了数据集之间的平衡力，并不断推动各类数据集朝着具有一定关联的方向整合、融合、聚合，这种强活性就是聚合力。聚合力强调的是解构原有数据平衡系统，并重构数据平衡系统，从而实现数据价值的提升与再造。聚合力可以用聚合力指数进行测度，主要用于度量数据流在完成块上集聚、关联之后挖掘出的新数据的价值密度。

上述三个环节中，平台化是基础环节，关联度是关键环节，聚合力是核心环节。第一，平台化是块数据形成的基础性条件，强调的是数据流动与聚集，对数据进入关联与聚合环节发挥重要作用。第二，关联度是块数据形成的关键环节，关联度是在数据流动与聚集的基础上，强化

数据之间的显性连接，同时发现数据之间的隐性连接。关联度是建立连接、剔除冗余及发现价值的过程，特别是数据隐性价值的发现过程。如果数据只是流动与聚集，未建立连接与发现价值的话，那么就无法为聚合力的实现提供基础，进而影响块数据的形成。第三，聚合力是块数据形成的核心环节，在这个环节要完成数据价值再造，而实现数据价值再造是块数据具备核心竞争优势的重要条件。聚合力是从解构到重构的过程，这个过程首先要打破在关联环节建立的数据平衡，然后通过新一轮的建立连接与发现价值而完成价值再造。通过三环节的关系解析可知，块数据形成的过程就是数据流动和聚集，连接建立及价值发现，关联建立和价值再造的过程，并且这个过程是循环往复、螺旋上升的。

第二节 平台化：从数据到数聚

（一）从平台到平台化

从平台到平台化，描述的是海量数据通过流动实现聚集的过程，也就是从数据到数聚的过程。在这个过程中，孤立的数据得以聚集，为发现及挖掘数据价值奠定了基础。

平台通常表现为一种空间、场域或载体。块数据平台是一种创新性和综合服务性平台，其表现形式多种多样，既可以是虚拟的，也可以是具象的；既可以是一种空间、一种场域，也可以是一种载体。第一，块数据平台为各主体投放数据、分析数据、应用数据提供了空间环境或条件。这种空间可以有边界，也可以无边界。平台通过在空间中建立相应的机制引导各主体进入平台，从而满足各主体对数据的需求，实现数据价值的最大化。第二，在块数据平台上会形成巨大的数据连接网络，这些网络会形成不同的数据场域。数据场域具有相对独立性的空间，相对独立性是不同数据场域相互区别的标志，也是不同数据场域得以存在的依据，不同场域之间既相互独立又相互关联。第三，数据的聚集、关联和聚合的过程都需要支撑载体，为其提供物质基础。从聚合角度而言，块数据平台是吸附数据的载体，也是全息化重构的重要载体，主要表现为经济社会被数字化离散表达之后，需要有一个空间来承载或支撑其重构的过程。

平台化是数据流动与聚集的动态过程。从平台到平台化描述的是数

据流动的过程，这种流动不仅能够促进数据聚集，也能促进数据社会的发展。第一，平台化是数据流动的动态过程。平台化的本质是数据的流动。平台为数据的自由流动提供了一种可能，这种流动实现了数据之间的连接。第二，数据流动促进数据聚集。在块数据中，平台化强调数据流动的速度，数据流动越快，数据之间建立的连接就会越多，就越能促进数据聚集。或者说，流动数据占全样本数据的比例越大，从数据到数聚的过程就越有可能建立连接并发现价值。就如同人们之间的交往一样，认识的人越多，就能与越多的人建立连接，从而接触到更多的人际圈，也就增加了融入更多人际圈的可能。第三，数据流动有利于促进数据化社会发展。随着大数据技术的发展，现代社会逐渐成为一种数据化社会。流动的数据可以消除数据的不对称，破除时间和空间的限制，实现数据之间的更多碰撞，发现并重构价值，从而促进数据化社会发展。换言之，数据之于数据社会，就如同水之于城市一样，江河的流动滋养万物生长，水一旦停止流动就会变成一潭死水，不仅影响自身活性，也会影响周边的生态系统。

（二）数据迁移

信息技术的发展使原有的信息系统不断被功能更为强大的新系统所取代，数据迁移是将仍然需要的历史数据转到新系统的过程，其主要特点表现为需要在短时间内完成大批量数据的抽取、清洗与装载。在块数据的语境下，数据迁移是将数据从一种数据环境移入另一种数据环境的过程。数据迁移的步骤通常包括识别源数据、确定数据集成转换的规则、转换质量评价。首先，块数据平台会对进入的源数据进行识别，初步确定数据的可用性；其次，会确定数据集成转换的规则，例如平台预先指定教育数据要与收入数据聚集，那么二者在平台上的所有数据便会根据规则进行聚集；最后，数据迁移可进行质量评价。数据迁移的成败与质量好坏影响新系统能否成功启动。如果数据迁移失败，新系统将无法正常启用；如果数据迁移的质量较差，还存在垃圾数据，新系统将面临很大的隐患，它表现在新系统一旦访问到垃圾数据，将会产生新的错误数据，甚至导致系统异常。

数据迁移是由系统数据量的不平衡性引起的，具有不平衡机制、逐利机制和趋同机制。

数据迁移的不平衡机制。相对于传统数据迁移的单向流动而言，

在块数据平台上，数据量若分布不平衡，数据会在数据密集区域与稀疏区域进行双向流动，以发掘更多的连接与价值。数据迁移与人口迁移有相似之处，都是由于系统的不平衡性发生的流动，这是寻找机遇的过程，也即寻找更多连接的可能性。新古典经济学派认为国际人口迁移是由全球劳动力供需分布的不平衡所引起的劳动力调整过程，迁移被视为一种人力资本投资。注

数据迁移的逐利机制。新家庭经济迁移学派认为，迁移决策往往要依据全家人收益最大化的原则。c利益驱动是数据迁移的重要动力，就如同在人口迁移的过程中，人们总是趋向于流向资源更丰富和机遇更多的地方，块内数据也会向能产生更大价值的地方进行流动和聚集。数据迁移依据相关主体的价值需求进行流动，流动的过程遵循价值关联规律，一方面实现相关主体的价值需求，另一方面在迁移过程中数据本身也可实现价值提升。此外，数据迁移也遵循收益递增法则，数据规模越大，对新数据的吸引力越强，从而让它的规模变得更大，也更具吸引力，如此不断往复。

数据迁移的趋同机制。人口迁移网络理论认为，当大量迁移者在迁入国定居，迁移者网络可能形成，并给其他跨境迁移者的迁移行为施加影响，形成迁移惯性，促使迁移不断进行和规模更大。注数据迁移与人口迁移相似，数据流动过程中会发生聚集从而形成数据集，而数据集会吸附其他的同类型或同价值数据向自身靠拢，形成新的数据集，这个过程是持续进行的，其价值也会不断增长。这类似于黑洞效应，即数据质量越来越大，数据流转速度越来越快，便可以把周边相关数据全部吸附进去。注此外，数据迁移会受到经验关联的影响，经验关联可以理解为既往的数据流动路径。这个路径可以是系统预设定的，也可以是数据流动过程中新形成的。例如系统设定将健康数据与收入数据进行关联，这就形成一种数据经验关联，在以后的数据迁移中，相关数据就会按照这条路径进行迁移。

（三）平台生态圈

数据迁移的结果会形成新的数据生态系统，称为平台生态圈。在讲平台生态圈前，有必要先讲一下自然生态系统，以桑基鱼塘为例。在桑基鱼塘生态系统中，通过塘基植桑、桑叶养蚕、蚕茧缫丝、蚕沙与缫丝废水养鱼、鱼粪等肥桑形成循环的能量流系统，实现种桑、养蚕、养鱼

的共生发展。整个过程体现出了生物圈的包容性强、主体多元、新陈代谢、循环可持续发展等特征。

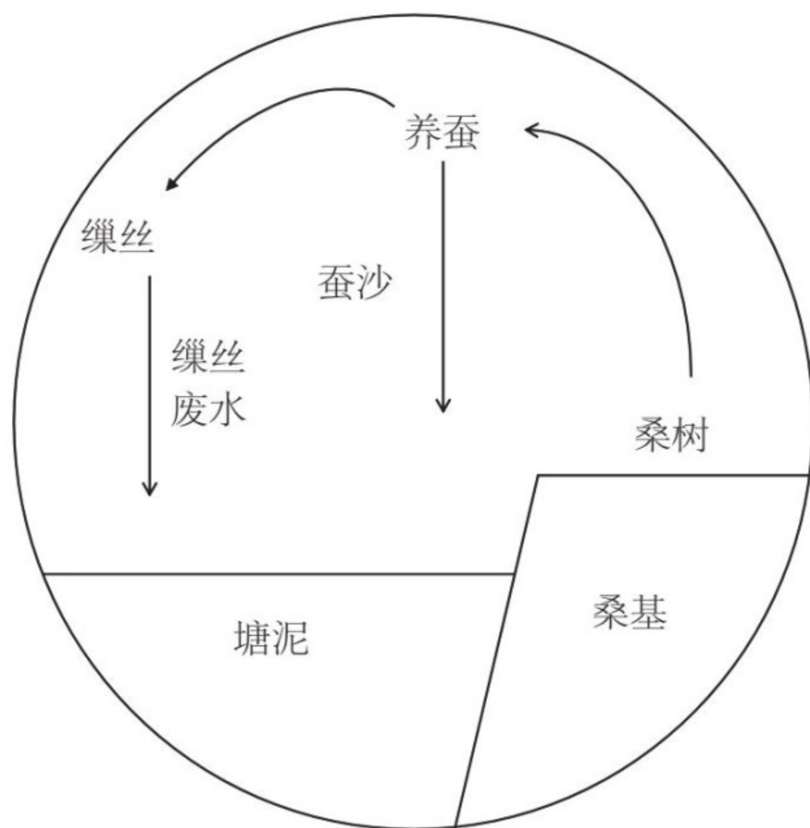


图4–2 桑基鱼塘示意图

平台生态圈与生物圈类似，具有自身的运行规则、包容性强、主体类型多元化、冗余剔除、可持续发展等特点。

运行规则。平台生态圈拥有自己的运行规则，这种规则不是限制数据主体的进入，而是促进数据主体之间相互作用。

包容性强。当数据主体进入平台生态圈时，具有不确定性，但平台生态圈都可以全部接受，无论其有无作用、价值量的大小如何、是否稳定、时效长短如何……只是平台会对进入后的数据进行清洗与冗余剔除，确保数据在每次聚集中的有效性。

主体类型多元化。由于平台生态圈的强包容度，平台生态圈里的

数据主体可以是数据集，也可以是单个数据；可以是个人数据，也可以是法人数据、社会数据、国家数据；可以是结构化数据、抽象化数据，也可以是暗数据。由此也带来数据价值多元化。

冗余剔除。剔除冗余是平台生态圈的固有机制。当数据主体之间相互作用时，那些未被利用的数据可能会被剔除，而这部分被剔除数据仍可以被循环利用，谋求与其他数据聚集的机会。

可持续发展。在平台生态圈内数据不断进行关联和聚合的过程中，数据会不断衍生出新的数据，而新的数据又会进行新一轮的流动与聚集，从而维持整个生态系统的平衡，实现平台生态圈的可持续发展。

第三节 关联度：从跨界到融合

（一）从融合到融合化

当今世界，各领域、各行业都在谋求互动发展与深度交融，以寻求更多的价值增长点，实现共生共赢。例如，传统产业与互联网的跨界融合正在加速，这种融合不仅可以促进电子商务市场高速发展，同时也让商贸、商超、零售业等传统企业实现转型发展。在世界融合发展的趋势下，数据的融合成为必然趋势，体现在物理区域及跨行业领域、跨数据类型的分类上。从物理区域来看，数据存储的硬件或者平台之间正在不断融合，比如手机、平板笔记本、电脑等智能设备通过接入数据中心实现融合。从行业类型来看，同类行业如网上订餐系统大众点评与美团外卖合并；不同行业如电商领域的阿里巴巴收购影视业的优酷土豆。从数据类型来看，个人数据、法人数据、社会数据与国家数据彼此之间可以互相融合。例如，企业生产数据与个人习惯数据相结合，产生了个性化定制与按需生产。

数据融合有6个基本步骤，它们分别是连接所需多源数据库并获取相关数据、研究和理解所获得的数据、梳理和清理数据、数据转化和建立结构、数据组合、建立分析数据集。注数据融合的瓶颈是思维。例如，水泥制作现在可以通过机械合成代替人工搅拌，但是石、沙、混凝土的比例与化学成分的添加等仍然需要人脑思维，目前机器不具有人脑思维的自发性和创造性。但是，随着块数据技术的发展，我们可以通过

大量数据的关联，为打破数据融合瓶颈提供解决方案。就如同IBM研究院给“沃森”计算机录入了大量的烹饪数据库，其中包括网上菜谱、美国农业部营养成分表以及有关美味菜品的研究报告。凭借这些数据，“沃森”计算机从食材资料以及现有的菜式资料中创制出新菜品。注

数据之间相互关联才有意义。如果关联程度低，平台化形成的数据之间或数据集内部就会呈现出黏性低的特点，这会导致数据之间关系割裂或断裂，通常表现为数据结构碎片化、数据逻辑混乱化、数据主体孤立化，进而数据的含金量减少，数据的可挖掘程度降低，融合化的难度加大、有效性降低。

融合化是深度关联的过程。基于块数据高度关联这一基本属性，数据的深度关联是数据融合的基本问题，各类数据内在的、紧密的持续关联是实现数据融合的基础。融合让不同领域、不同类型的数据成为高度关联的整体，最终形成新的数据形态。

融合化的关键是打破数据壁垒，实现跨界关联。传统的社会分工使不同领域和行业的数据以“条”的形式存在，加之行业因素、利己主义及数据风险防范等原因，造成数据壁垒，成为数据融合的障碍。只有打破数据壁垒，才能实现数据之间的自由流动，实现跨界关联以及融合，最终形成新的关联模式。

（二）数据群聚

群聚效应原是一个社会动力学的概念，用来描述在一个社会系统里，某件事情的存在已达到足够的动量，使它能够自我维持，并为以后的成长提供动力。例如，企业经过专业化发展，当主营业务壮大到一定程度，其上下游产业链的相关企业就会聚集到附近的地理区域，当聚集的企业涵盖的产业链足够完整、数量规模足够大，便形成大型企业集群。产业集群会为所在区域提供更强劲的竞争力，成为区域经济新的增长极。

数据群聚同样如此，当数据、数据集、数据关系的数量积累到一定程度的时候就会突破临界点，并持续地按照一定规则自由流动，最终实现自我发展、自我成长。例如，人工智能会随着人们的使用和数据量的不断输入实现自我改进，把在一个项目上习得的东西运用到其他项目上。就像给AlphaGo输入一定量的棋局算法数据，当这些数据累积到一

定程度并突破临界点时，它就能进行深度学习，进行自我对弈甚至自我开发棋局，同时通过与人对弈，完善自身的数据系统。

数据群聚的前提是去边界。去边界就是打破时间的边界、打破空间的边界、打破现实与虚拟的边界、打破数据或数据集之间的边界。去边界是一种力量，这种力量让一切毫无关联的数据与数据集都相互关联起来，完成从跨界到融合的过程。关联的数据或数据集具有社交性，表现为流动与开放。边界的消融会带来数据社会的进化。

数据群聚具有自发生长机制与互惠共生机制。一方面，从单个数据群落来看，通过借助数据群聚效应能使自身数据量、数据节点自发生长，数据结构自我优化，实现数据群落有机体的良性发育。另一方面，每个数据群落通过互惠、竞争等相互作用而形成巧妙的组合。从相互间的依存关系而言，这种组合存在寄生、偏利共生、互利共生等复杂情况。寄生是指数据从一个系统脱离转入另一个系统；偏利共生是指数据间相互作用对一方没有影响，但是对另一方有益；互利共生是指数据间相互作用对双方都有利，让相互作用的数据都产生更大的价值。

数据群聚形成的数据群落之间不是随意地拼凑在一起，而是有规律地组合在一起，这样才能形成稳定的群聚状态。首先，数据群聚要遵循量的原则，也就是说数据聚集要达到一定的量才能产生群聚效应，才能呈现出一种互惠共生的状态。其次，数据群聚遵循结构规律，数据群聚是有结构的，其结构越复杂，数据就越有可能被多层面、多维度地利用，数据系统中数据价值的利用就越充分，从而逐渐形成一种相对稳定的群聚结构。再次，数据群聚必须遵循存在条件规律，就像生物群落需要温度、湿度、土壤、高度等环境条件一样，数据群聚的形成需要基础设施、技术条件、开放程度、价值需求等存在条件，这些条件越适宜数据群聚才越容易实现，发挥的作用也越大。

（三）“大数据×”效应

继“互联网＋”后，“大数据×”的概念也逐渐为人们所认同。这种充满聚合效应、乘法效应、外部效应的内在解构变化产生大量难以预估的关联反应，印证了凯恩斯经济学理论中最为基本的乘数概念。

“大数据×”的乘数效应是数据或数据集之间通过关联对数据生态系统中的其他数据或数据集产生的带动作用，通过循环累积关联从而不断

地强化和放大影响。在乘数效应中，一个数据的变化将会使整个数据生态系统产生连锁反应。与加法效应相比较，或许能更好地理解乘数效应。从关系的着眼点看，加法效应强调“要素”的完整性，乘法效应更强调要素之间关联所产生的效能；从叠加的效果看，随着次数的增加，加法效应中每一次努力带来的成果所占的比重都越来越小，而乘法效应则会出现几何级数增长的形态。[注]乘数效应会带来数据量、数据集、数据关系、数据融合速率等的倍增，从而拓展新的领域，撬动大能量，实现指数级增长。

“大数据×”的运行机制是数据涟漪。数据涟漪是指每一个数据节点的变化会影响关联数据节点的变化，进而影响二次关联数据节点的变化，催生新的多种反应，这种新变化又会经过多次关联反作用于最初的变化节点，致使这个节点又产生更多新变化。在经历过融合化的关联过程后，海量数据编织成一张紧密关联的数据大网。这张无边界无间隙的大网上每一个节点的变化，都会泛起一阵数据涟漪，这种反应是循环往复、层层递进、辐射扩散的交叉影响。数据涟漪通常会表现为数据之间的交叉融合，交叉是历经融合及融合化之后边界的消融。交叉最主要的表现形式是数据集交叉和行业交叉。例如，“大数据×”让政府的数据权力边界在原有的管理行业和产业的方式上面临一次重大的交叉重组和调整。

“大数据×”产生的效应是由“小”决定未来。数据或数据集通过乘数效应形成庞大而精确的数据网，这个数据网会呈现出如下4种状态。一是小而美，可以让未来组织结构变得精简，从而节约资源成本，找到资源配置的最优方案。二是小且精，它打破传统建模的单向孤立计算，通过计算数据间的周期交叉影响，模拟多点位波动的现实环境，达到精准化预测。例如亚马逊通过分析顾客的消费记录、搜索记录、鼠标在某个商品页面上停留的时间等数据实现预估发货。三是小见大，某一领域数据网络中的数据量越大，就会越接近并越能描绘现实世界，这运用于决策支持系统，能找出实践活动的最佳解决方案。四是小及远，可以应用于安全监测和风险防控系统，在负面事件尚未发生之前，先行解决隐患。例如，通过分析危险品的购买主体背景、购买数量等数据，实现对风险事件的预测与预防。

第四节 聚合力：从解构到重构

（一）从平衡力到连接力

平衡具有自我恢复机制。平衡力是指数据生态系统在没有外力的刺激下会不断进行自我修复并处于平衡状态的过程。即使内部由于某种原因发生更迭，也能很快进行自我修复、重归平衡，这种平衡没有改变原先平衡的性质。数据这种缺乏外力刺激且处在自我恢复的平衡状态中的特点，会使整个数据系统缺乏活力与创造性。

数据平衡影响价值增长。当数据处在平衡状态时，其流动性与活性会降低，一方面会影响数据价值增长，另一方面则会影响数据集与其他数据集之间新一轮的连接建立与价值发现。这类似于公司的成长轨迹，当公司发展成熟，内部系统趋于稳定平衡时，也会产生缺乏活力与创新性等问题，并对外部问题的“免疫力”下降，这会影响公司的整体运行发展，甚至造成公司解体。

连接力打破平衡，重新发现价值。连接力打破数据平衡，在这个过程中不断破旧立新创造新的数据平衡。数据平衡打破不是平衡消失，而是通过打破旧系统的平衡，创造出新的数据平衡系统，从而产生新的数据价值，正如克莱顿·克里斯坦森[注]所说的，通过破坏找到一种新的路径、新的生产函数与模式。

（二）数据引力

通过数据引力波，数据在数据引力场中聚合、裂变，这个过程就是数据引力发挥作用的过程，也是不断打破数据平衡、形成数据连接、再造数据价值的过程。

数据引力是打破数据平衡力、建立连接力的内生动力。数据引力推动数据空间[注]中的数据或数据集之间发生不规则、无边界、多维度和复合叠加的运动，它们相互吸引、共同作用，如此循环往复。一方面，吸引其他数据聚集，实现量的积累；另一方面，“数据产生数据”引发质的变化。整个过程中，数据引力让数据不断聚合，当数据聚合到一定量的时候，会突破临界点打破数据原有的平衡力，建立连接力，产生新的价值关联。这就好比核聚变的过程，氘和氚在极高的温度和压力下让核外电子摆脱原子核的束缚，让两个原子核互相吸引与碰撞，产生聚合作用，生成新的、质量更重的原子核，中子和电子在碰撞中被释放出来，产生巨大的能量。

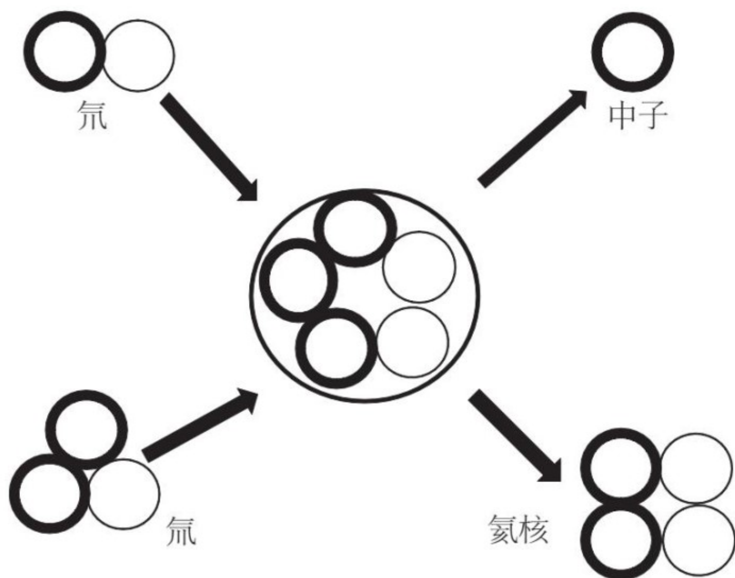


图4–3 核聚变示意图

数据引力的作用不仅存在于数据与数据之间，也存在于数据与数据空间、数据空间与数据空间之间。数据引力发生作用的机制主要有如下三种方式。第一，将其他数据或数据空间变成自身有机的组成部分，实现系统再造。就如同植物嫁接，将植物体的一部分固定在另外一个植物体上，使其组织相互愈合，培养为独立个体，嫁接母体利用砧木的有利特性，具备了早结果、高抗寒性、抗旱性、抗病虫害、产量高等新能力。第二，数据在引力的作用下被分解，原有的形态被打破，数据质点重新排列组合，产生新的数据或数据空间。这个过程类似于化学反应中高分子形成的过程，单体小分子通过相互链接成为链状大分子，尤其是非常大的分子，最后形成高分子，高分子就是聚合的产物。第三，数据价值密度决定数据引力的强弱，数据引力的强弱改变数据之间的组合方式，进而造成数据空间的形态变化。比如当水分子之间的引力最强时，就形成固态的冰；当其受到温度或压强的影响，分子间的引力减弱进而重新排列，则形成液态的水；当其引力最弱时，则气化成水蒸气的形态。

在数据引力的作用下，一方面使数据空间中的任何两个数据之间都存在关联，构建起多维、立体的数据连接网络；另一方面在打破数据平衡之后，能建立起更多新的数据连接，从而使原有的数据连接网络得到不断扩充、发展和完善，数据连接网络的建立与扩充最终实现数据价值

的再造。根据主体需求、数据质量、数据连接等的不同，再造的价值可以实现多元化应用。正如凯文·凯利所说的，从人们睡醒睁眼的那一刻开始，网络就试图预测人们的意图。在记录了日常生活数据后，网络会试图先人们一步，在人们提问前就给出答案；在开会前就给出文件；在和朋友吃饭前根据天气状况、地理位置、本周饮食情况、上次与朋友一起吃了什么等数据推荐出完美的就餐地点。[注]

（三）离散化解构和全息化重构

解构与重构是块数据高级形态的重要特征。离散化是对连续性的否定，是把连续型数据切分为不连续的若干段。离散化解构就是打破原来的数据结构，将其分解为不连续的、基本的数据单位，即数据元。全息化重构是对数据元多维度、多方位、多类型的重组。在平台化和关联度环节也可能会存在数据结构的打破，但这种打破是低级的、微小的，不会发生质变。就像拼图一样，虽然将图片剪成不规则的小碎片，但是这些碎片重新组合起来之后仍然是原来的那幅图，性质并未发生改变。在聚合力环节发生的离散化解构，这种打破是本质的、深度的，数据价值会发生质变。就像化学反应一样，分子在破裂为原子之后通过重新排列组合生成了新的物质。例如，盐酸与铁在反应过程中，盐酸被解构为氯元素与氢元素，通过化学反应，生成了新的物质——氯化亚铁与氢气。解构与重构强调数据价值产生之后，还可以再次返回块数据平台，进行循环的解构与重构，从而不断再造更高阶的数据价值，离散化解构与全息化重构过程贯穿于聚合环节的始终。

离散化解构是全息化重构的基础。离散化解构为全息化重构提供了可能，离散化的数字化解构过程不是终点，只是推动价值再造的第一步。与离散化解构过程相比，全息化重构同样相当重要，全息化重构是价值再造必不可少的关键环节。如果没有全息化重构，那么离散化解构就失去了价值[注]，全息化重构是离散化解构的最终价值体现。在离散化基础上，全息化重构的基本运行原理如下：数据平衡系统经过离散化解构之后，再经由数据处理平台进行全息化重构，形成一个全息的价值再造过程。之所以要对已有数据进行解构，并在此基础上进行重构，原因在于已有数据在解构之后通过重组可以产生更多的价值。正如纽约大学经济学家保罗·罗默所说，真正可持续的经济增长并非源于新资源的发现和利用，而是源于将已有的资源重新安排后使其产生更大的价值。圣塔菲研究所的经济学家布莱恩·亚瑟也说过，所有的新技术都源于已有技

术的组合[注]，增长来源于重组。

全息化重构的重点是价值再造。整个社会被离散化解构、进行数字化表达以后，还需要在价值基础上对离散化的数据进行全息化重构，通过全息化重构发现新的价值和商机。[注]法国后结构主义哲学家雅克·德里达指出，每一次解构都表现为结构的终端、分裂或解体，但是每一次解构的结果又都产生新的结构。数据、数据集、数据关系的解构与产品的解构相似，产品解构后的每一个“原件”，都会进行流动以寻找新的用途，并和新的产品绑定在一起。重构就是以主体需求为中心，对解构后的数据进行个性化改造，构成一个全新的、不同于以往的价值集合。从解构到重构是块数据的机制，每一次解构与重构的结果都会产生新的结构[注]，从而实现数据价值再造。

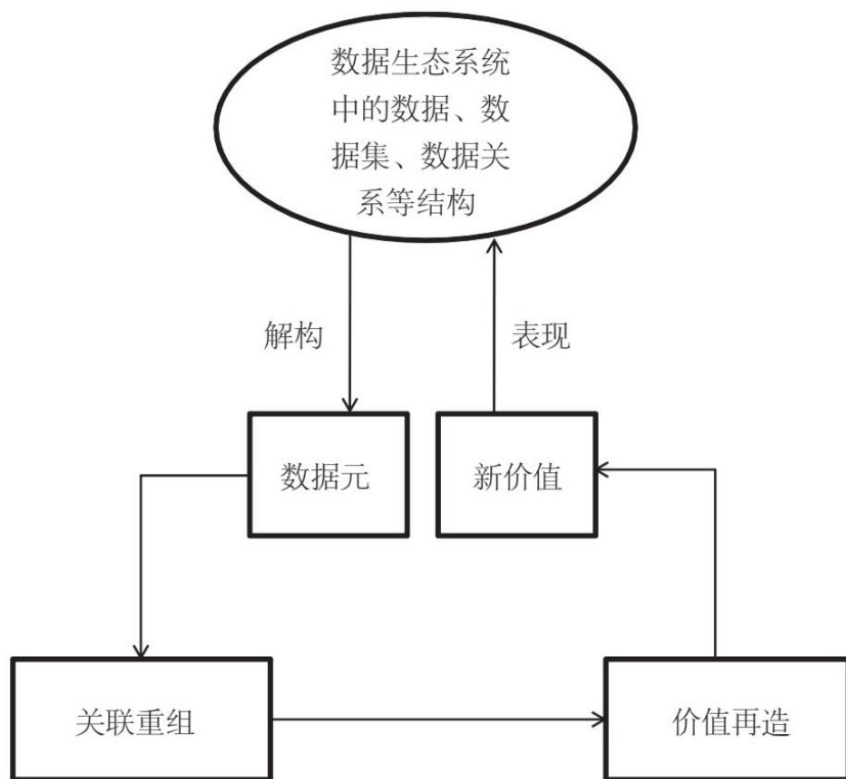


图4–4 离散化解构与全息化重构示意图

1. Hestenes, D. Modeling Games in the Newtonian World. Am. J. Phys, 1992, 60 (8), 732–748.

2. 王文清：科学教育中的建模理论，《科技信息》，2011年3月。

3. Halloun, I. A. Schematic Concepts for Schematic Modeling of Real World: The Newtonian Concept fo Force. Sci. Ed, 1998, 82, 239–263.

4. 图模型，http://baike.so.com/doc/6531519–6745255.html。

5. 陈玉：《D·H·劳伦斯对未知的探索》，四川大学，2004年4月。

6. 尼古拉斯·卡尔：《浅薄》，刘纯毅，译。北京：中信出版社，2015年。

7. 王婧韫：数据迁移的一般原则，《电脑开发与应用》，2000年4月。

8. 张晓青：国际人口迁移理论述评，《人口学刊》，2001年3月。

9. 张晓青：国际人口迁移理论述评，《人口学刊》，2001年3月。

10. 赵勇：《关注大数据就是关注活性与流动性》，财经网，2015年11月。

11. 张小彦：《数据融合：大数据分析的瓶颈》，财新网，2016年3月23日。

12. 凯文·凯利：《必然》，周峰、董理、金阳，译。北京：电子工业出版社，2016年。

13. 连玉明：《DT时代：从“互联网＋”到“大数据×”》，北京：中信出版社，2015年。

14. 克莱顿·克里斯坦森教授是美国哈佛大学商学院创新理论大师，它在其著作《创新者的窘境》（该书于2014年由中信出版社出版）中提出破坏性创新，改进了约瑟夫·熊彼特的创新理论。

15. 数据空间即前文所说的“数据引力场”中数据聚集后形成的形态。

16. 凯文·凯利：《必然》，周峰、董理、金阳，译。北京：电子工业出版社，2016年。

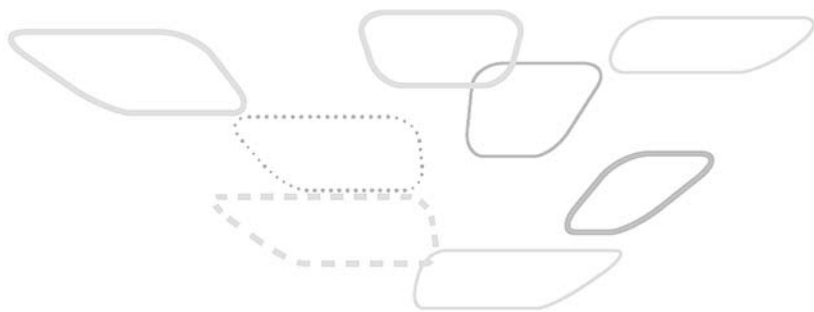
17. 徐晋：《大数据经济学》，上海：上海交通大学出版社，2014年。

18. 凯文·凯利：《必然》，周峰、董理、金阳，译。北京：电子工业出版社，2016年。

19. 徐晋：《大数据经济学》，上海：上海交通大学出版社，2014年。

20. 张静：《专家解读块数据：从解构到重构，从多维到共享》，中国科技网，2015年4月20日。

第五章
激活数据学



如果说块数据模型的提出是块数据在理论构建方面的重要进展，那么，块数据的应用则是一个更具价值的研究领域。大数据时代是一个更加开放、更加复杂的巨系统，对不确定性和不可预知性实现更加精准的预测，是块数据实践应用亟待突破的关键性问题。正是在这一背景下，贵阳的决策者和管理者创造性地提出了激活数据学的理论和实践体系，并在一些重要领域开始了系统性探索。围绕贵阳的思考与实践，我们进一步从理论上对激活数据学进行研究，探索其理论基础和运行规律，并且尝试通过量化手段进行模型构建。

激活数据学的基础是人工智能的飞速发展，通过人机交互推动高度数据化的智能与高度智能化的数据相互融合，对高度关联的数据进行碰撞与激活，进而实现对不确定性和不可预知性的精准把控。激活数据学将颠覆并替代传统的思维范式，集合多领域跨学科的系统，基于复杂理论在大数据领域开辟一个新的战略制高点。

第一节 复杂理论与激活数据学

（一）复杂理论：第三代系统思想

英国著名理论物理学家霍金认为“21世纪将是复杂性的世纪”，作为继还原论、系统论之后的“第三代系统思想”——复杂理论，正是解决这

类问题的前沿理论。

第一代系统思想是由蒯因提出的还原论。1951年，还原论首次出现于美国逻辑哲学家蒯因的《经验论的两个教条》中。还原论是一种主张把复杂高级运动形式还原为简单低级运动形式的哲学观点，该观点认为整体是由个体的简单相加或机械组合而成的。还原论的主要不足是其忽略了系统内部的关联属性，而将其定义为一个可以分割的简单复合体，这种接近于线性的理论思想根本无法支撑起对复杂系统的研究。所以，在经历了长期的实际验证之后，还原论已经被认为是一种不切实际的理想型理论，并逐渐被第二代系统思想——系统论所替代。

1948年，美籍奥地利生物学家贝塔朗菲创立了第二代系统思想——系统论。系统论是强调系统整体观念的研究理论，该理论认为任何系统都是一个不可分割的“有机体”。系统论主张针对系统内在具有的复杂关联性进行研究，并且提出了“整体功能大于个体功能之和”的重要非线性概念。然而，系统论的主要不足之处在于，没有深入发现系统内主体的适应性以及其在自适应过程中产生多层次复杂功能结构的能力。

在复杂理论兴起之前，对具有高度复杂性、多样性和无序性的个体因素及其自组织现象的忽视和排除，是长期占据统治地位的经典科学研究方法的不足。为了弥补这些不足，复杂理论的相关研究应运而生。1973年，法国当代著名思想家埃德加·莫兰提出的“复杂性方法”是复杂理论进化过程的开端，他纠正和批判了经典科学的还原论和机械论，主张用“有序与无序统一作用、整体与部分共同决定”的“复杂性范式”去认识和决定系统。1979年，比利时学者普利高津在其基础上确立了“复杂性科学”这一概念，不仅揭示出物质的进化机制具有耗散性，而且得出了“系统变化是使动态有序结构不断增长和多样化的不可逆过程”的重要结论。1984年，圣菲研究所在“复杂性科学”的基础上开始了对“复杂适应系统”的研究，不仅得出“适应性造就复杂性”的研究结论，而且描述出“复杂适应系统都具有自组织的能动机制的共性”。1994年，在之前研究成果的基础上，约翰·霍兰德成功继续了对复杂理论的完善，即确立了“复杂适应系统理论”。

复杂理论弥补了两代系统思想的不足。复杂理论首先承认了个体因素的重要性，又主张用“有序与无序、整体与局部协同作用”的复杂范式对各种系统展开研究。不仅揭示了系统的“非线性与非平衡”的重要研究价值，而且为系统研究领域明确了一个重要的指导思想：主体的适应性

造就系统的复杂性是所有复杂系统内部的隐秩序。

作为系统解释块数据微观个体和宏观系统的理论支撑——激活数据学，对复杂理论进行了有益的继承和延续。根据激活数据学，块数据具有在非平衡和非线性共同作用下的高度灵敏性，并在宏观层面表现出长程的秩序并演化出多样化的自激活、自流程、自组织状态。块数据一旦被激活，其所蕴含的数据价值被释放和放大，并且将产生大于个体价值总和的放大效应，达到“$1+1>2$”的效果。

（二）引爆点：沙堆实验的启示

沙堆实验是复杂理论的代表实验。沙堆实验的研究过程是让沙一粒粒下落到平面上，沙粒会自组织升高成一个圆锥形沙堆。沙粒下落所带来的冲击力与沙粒间原有的引力进行碰撞所产生的能量可以贯穿整个沙堆。这种能量不仅可以破坏局部沙堆的平衡，还会减弱沙堆的整体弹性及恢复力，使沙堆逐渐趋近于“自组织临界点”。沙堆到达临界状态以后，所有的沙粒都会处于具有高度灵敏性的一体化接触中，这时，任何单一沙粒的下落都可能引爆整体的崩溃。这个引爆沙堆从量变到质变的“临界沙粒”就是引爆点。而引爆点出现的前提，是沙堆趋向崩溃的过程等同于一个复杂自适应系统产生高度灵敏性的过程。

沙堆系统的非线性，决定了引爆点的不可预测。在沙堆整体崩溃之前，已经发生过多次的局部坍塌，局部的坍塌虽然不能立刻导致整个系统的崩溃，却是对整体系统弹性和恢复力的不可逆耗散。耗散是一个不可控的过程，会造成沙堆整体引爆的非线性结果，也决定了引爆点的不确定性。

沙堆系统的非平衡性，是一种自组织临界性。需要注意的是，这种自组织临界性不同于平衡态统计力学中所指的平衡相变的临界性。平衡系统的状态转化过程是通过调节系统的某个参数而达到的。然而自组织临界性的产生不需要依赖外界因素来干扰或驱动，它是通过系统内部的各组成单元之间的相互作用产生，纯粹是系统自身的一种动力学机制。

类似于沙堆构成的系统，许多复杂系统的运行状态都可分为亚临界、临界和超临界三种。在正常环境下，系统会自发地朝着临界状态进化，然而一旦发生突变情况，系统可能进入超临界状态并爆发“崩塌”现象。令许多生物学家困惑的“寒武纪生命大爆发”，很有可能就是自组织

临界性的机制所产生的。它不仅是一次地球生命进化过程中的爆发式事件，更有可能是生命进化过程中一个自组织事件，因为这一事件具有极明显的自发性进化行为。

沙堆实验所揭示的系统非线性和非平衡性，是构成激活数据学系统特征的最基本也是最重要的影响因素。

（三）网络、扩散和反馈

从众多的复杂系统研究成果看，大都包含有许多基本个体和基本个体互相作用两部分，这两个基本部分构成的模型对复杂系统具有最广泛的意义。网络、扩散和反馈是激活数据学观察系统个体间相互关系与影响的三个重要概念。其中，网络刻画了个体关系的基本结构，是激活数据学的基础模型；扩散是单个个体与其他个体接触、连接、融合和决策的动态互动过程；反馈描述的是动态网络演化中网络结构与网络动力学之间的双向影响。

网络。在激活数据学中，把复杂系统中的基本数据单元看成结点，把数据单元之间的相互作用看成结点之间的连线，而把网络看作激活数据学的一个最基本的数学模型。从某种意义上说，网络更多的是一种思考问题的方式，有利于更好地理解不同个体之间或者复杂系统之间的相互作用。例如，大量的研究关注于网络对各类现象的影响，包括创新的扩散、金融市场中的信息流、全球价值网络、新媒体传播等。这些研究均认为，个体之间相互作用的网络拓扑结构对于理解系统总体现象和个体自身行为特征具有重要的作用，比如静态结构网络上的扩散过程，或网络结构改变之后对个体行为的影响等。

扩散。扩散是复杂系统理论中的重要研究命题，涉及产品、技术、思想、行为、策略以及一些经济社会现象等在系统中的传播。激活数据学所研究的扩散，其实质是达到一定数量的数据个体感知并适应外部环境变化，并与其他个体接触、连接、融合和决策的动态互动过程，它对于人们理解和控制块数据系统具有重要意义。

对激活数据学中扩散规律的研究，从早期的宏观扩散视角（研究整体扩散效果）转向微观扩散视角（研究扩散的微观基础），其主要原因在于系统中个体决策过程是个体间相互影响、共同决策的累加结果，换句话说，扩散的微观基础在于个体间的互动过程，这是决定扩散的根本

原因。在继承传统扩散理论和方法的基础上，激活数据学表现出如下一些新的特征。第一，个体间相互作用可以影响扩散的结果。网络中各结点间存在相互作用，网络中其他个体的选择会形成、影响甚至改变个体决策和行动，进而改变扩散的进程和结果。第二，网络效应对扩散过程具有重要影响。网络效应是指网络结点的行为改变或决策收益不仅取决于初始信号本身，而且受到涉及个体数量增减的影响。例如，电话、网站、国际标准竞争、新媒体通信工具、电子文档文件格式等，都是具有“网络效应”特征的商品——参与使用的人越多，带动其他新用户跟进使用的吸引力也就越大。不过需要注意的是，网络效应分为正效应和负效应，当个体决策收益随着新近加入的其他个体数量的增加而增加时，网络存在正效应（或策略互补性等）；反之，当个体决策收益随着新近加入的其他个体数量的增加而减少时，网络存在负效应（或策略替代性等）。从另一个角度来说，也可将这两类特征看作网络个体选择所体现的规律性，个体选择的趋同性，导致网络出现正效应；而个体选择的相异性，则会导致网络出现负效应。无论是网络效应，还是个体选择的规律性，都会显著影响网络扩散结果。第三，网络拓扑结构的差异可能导致扩散出现新的变化。例如，目前研究显现出的诸如小世界效应和无标度特征可能导致扩散过程出现新的变化。

反馈。在现有的研究中，较多关注的是网络结构对功能的影响，即在静态的网络结构上，网络结构如何对动力学过程产生影响。但事实上，网络总是处于不断变化的动态过程中，个体间的关系也时有改变。在不同时间尺度上都会有结点改变相互间的连接方式（如两个企业之间建立合作关系）。尽管从一般意义上来说，当系统规模足够大时，这些变化在某一时点显得微不足道，但从较大的时间尺度来看，这种微小变化的累积可能改变整个网络结构，并影响到网络上的扩散动力学。激活数据学认为，网络结构与网络上的动力学之间存在着双向反馈过程。在给定初始条件的网络中，自适应个体不仅可以更新自身策略，还可以根据收益情况选择连接个体，与之建立或中断互动关系，进而构造个体间的扩散过程与个体间网络结构双向反馈的共同作用进程，最终推动动态网络的演化。归纳其基本特征，个体行为与网络结构对扩散的影响，取决于扩散原点是否具备相对优势；系统整体价值不仅是个体收益的累加，也与所处的网络结构有关；个体行为会影响和改变网络结构特征，扩散能够改变网络结构的拓扑特征。

第二节 激活数据学的运行规律

（一）数据搜索

数据搜索是激活数据学中的准备阶段，是块数据系统依据某种信号组织相关数据的一种行为。激活数据学强调的是对所有关联数据的整合、分析和创造性挖掘，需要通过数据搜索来实现整个关联数据体系的建立，为数据处理提供尽可能完整的数据资源基础，以确保处理结果的准确性，同时防止出现数据价值挖掘的盲点。

激活数据学中的数据搜索延续了传统数据搜索的原理和各种技术，即将用户的需求在原始数据库中进行比对，通过匹配机制计算信息的相似度，并输出结果。例如搜索引擎就是对网页中的关键词进行索引，建立索引数据库的全文搜索引擎，当用户输入某个关键词时，搜索引擎将请求信息进行量化，与数据库进行对比，通过云计算将包含该关键词的所有网页都搜索出来，并按照与关键词的相关度高低依次排列。

激活数据学中的数据搜索有自身独特的优势。首先是搜索范围更全面，结果中会包含强关联、关联、弱关联以及潜在关联的数据，甚至还能包含当前并不关联，但在未来可能会发生关联的数据，它的智能程度达到了“走一步想十步”。其次是数据搜索的自发性，传统的数据搜索是被动的，都是需求在前、搜索在后，而激活数据学中的数据搜索会在分析前期关联的基础上，预见性地进行自主搜索，为更为精准的预判搜集全面的数据资源。

（二）关联融合

关联融合是激活数据学中的预处理阶段。它的目的不是将所有数据集中在一起并标准化而产生唯一的真相。它是以产生多元价值为目标将多种数据源中的相关数据提取、融合、梳理整合成一个分析数据集。

注这个融合结果是个独立和灵活的实体，可随数据源的变化重组、调整和更新。

激活数据学中的关联融合是一种跨界的融合，类似于“互联网＋”模式，如阿里巴巴创造的电商模式、苏宁的“亚马逊＋沃尔玛”模式，都是对行业间的条数据进行跨界融合。例如，线下零售行业在多年的运营过程中，积累了大量的用户订单数据，这部分看似孤立的数据若能与其他

行业的数据融合，将会形成多向的价值空间。

激活数据学中的关联融合更加强调各数据主体间更高阶的融合，以个人、企业、社会等为起点来形成关联，将数据置于特定背景中，探索数据现象和其他社会现象之间的关系，探索集合体中的内在秩序，并根据这种更为本源和微观的内在秩序建立关联并实现融合。例如，未来某天你在商场里试穿了一件外套，却由于价格问题最终没有购买。之后在网上你看到同样的外衣广告推送，当点击后你会惊喜地发现，网页显示的竟然就是你的尺码和你喜欢的颜色，甚至还有团购网站专门提供的100元优惠券！激活数据学中的关联融合更加系统和全面，根据深层、微观的数据关联融合，反馈作用于数据搜索，进行相关数据资源的汇集整合。为激活数据学中的自激活提供了不断优化的、动态的数据资源基础。

（三）自激活

自激活是激活数据学研究的核心环节，是数据价值释放的临界点。如图所示，经过数据搜索和关联融合的预处理阶段后，数据单元准备进入自激活阶段，这个阶段的实质就是一个筛选阶段，能够自激活的数据单元才能实现热点减量化和智能碰撞，释放价值潜能。

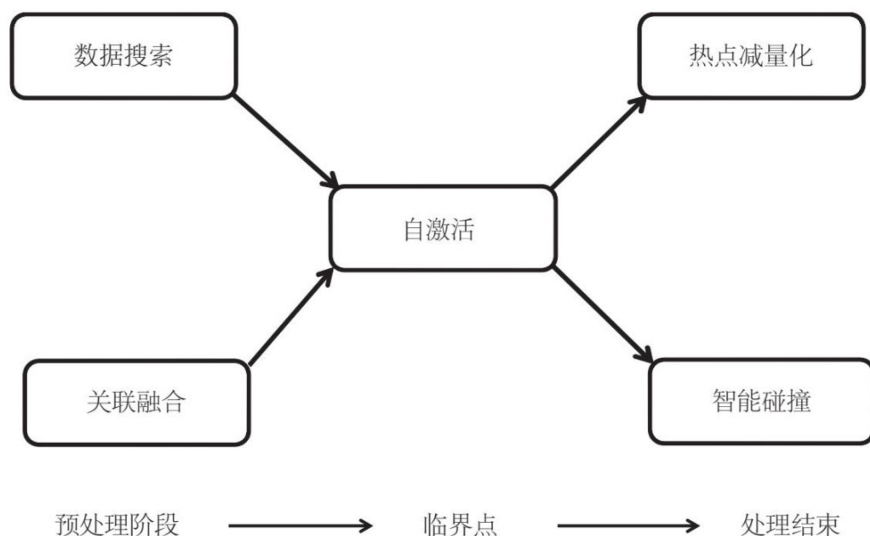


图5–1 激活数据学运行环节的划分

数据单元实现自激活的过程，类似于人类神经系统中神经元的活动情形。中枢神经系统由大量的神经元构成，仅是人的大脑皮层就有大约

1 000亿个种类繁多的神经元。神经元的兴奋状态和抑制状态是由所接受的信号强度来决定的，当信号强度达到一个临界值时，神经元就会进入兴奋状态，否则其仍然处于抑制状态。数据单元活跃程度呈现三种不同状态：

潜动源状态。即低维度的动源，此时数据单元处于休眠等待状态，只进行热点数据处理，如简单的存储和数据交换。就如同人处于睡眠状态一样，大脑并没有停止运转，虽然不会进行复杂的逻辑思考，但它仍然具有思维，是在活动的，而且可以随时被唤醒。处于潜动源状态的数据单元也可以被随时唤醒并进入自激活状态。

动源状态。动源状态是数据单元的一种常态，相比潜动源状态，数据单元在该状态下较为活跃，不仅会进行热点数据处理，还会根据一定的算法规则进行热点逻辑计算和分析预判。

热动源状态。当数据单元活跃度足够高时就会进入热动源状态。处于此状态的数据单元具有很高的能值和辐射力，它可执行热点数据计算、热点逻辑计算，还可以依靠自身的高活跃度带动临近的数据单元节点，一定程度上影响它们的状态跃迁。

在激活数据学中，数据单元的状态切换是靠对信号的响应来实现的。信号是不同数据引力波的组合形态，它可以来自系统外界，也可以在系统内不同主体间的互动过程中产生。系统外界刺激生成的相应信号，通过数据引力波在整个系统中进行传递。系统中的数据单元，尤其是处于热动源状态的数据单元间也可能会相互作用形成信号，通过相应频率的数据引力波进行扩散。

不同的数据单元对于信号的响应情况也存在不同情形。有的数据单元几乎不受信号影响，依然处于不活跃的潜动源状态；当信息强度稍强时，数据单元就会激活进入动源状态；当传递的信号足够强时，就可能发生“共振”现象，数据单元就会被激活并迅速进入非常活跃的热动源状态，形成巨大的能量。

（四）热点减量化

热点减量化是数据单元自激活后，在系统层面出现的帕累托最优状态。帕累托最优是指资源分配的一种理想状态，是公平与效率的“理想王

国”。热点减量化就是实现数据处理资源优化配置的理想路径，通过自激活步骤，对所有数据单元活跃状态进行清晰的层次划分，并以此为依据进行热点的过滤和筛选，实现热点的减量化，挑选出更具价值的数据单元进行分析，这在很大程度上提高了数据深度处理的效率。

热点减量化能够降低数据噪声，排除不准确、失去时效性、关联度不高的数据，使得保留下来的热点数据更精准、更有价值，保证最终数据处理的分析结果更加准确。例如，电商网站1号店后台的价格智能系统每天实时在线搜索60多个网站和1 700多万种商品的库存信息和价格信息，并根据竞争对手的商品价格实时调整自己的商品价格。但是一些季节性、节假日的数据，还有消费者的个人操作中包含的不少无效行为数据等都属于噪声数据，会对预判结果造成干扰，这些噪声数据都会被过滤掉。

热点减量化能够提高数据的处理效率，实现资源的最优配置。很多数据科学家、数据分析师、数据库管理员都认为，项目的80%的时间和经费花在了数据的准备工作上。其中多源数据的处理是最耗费资源的任务之一。通过热点减量化就能够打破这一困局，实现数据计算处理资源的合理分配，将处理能力集中于具有高价值的活跃数据单元上，提高数据处理的效率。

热点减量化还代表了数据处理从繁到简的必然趋势。在大数据出现之前，人们所从事的都是数据的增量分析，而随着数据生产规模的剧增，关联节点增多并日趋复杂，从数据噪声中提取价值数据的难度增大，处理结果的准确性和价值密度也日渐降低。如果不改变数据增量分析的路径，将会在数据处理上倾注更多的成本，逐渐拉大投入和产出之间的差距。因此，采取减量化的数据分析路径，才能在有限成本约束的条件下实现数据价值挖掘的最大化。

（五）智能碰撞

智能碰撞是数据单元被激活后，在宏观层面上涌现出的数据价值创造和放大的过程。自激活进入热动源状态的数据单元，利用彼此间的高度活跃性互相融合聚变形成创新性的信息，并释放大量的数据价值。激活数据学中的智能碰撞类似于人类的头脑风暴，群体中的每一个人都根据自己的自由联想和思维逻辑表达出相关的想法，在相互启发、联想和碰撞过程中激发出创新设想。

智能碰撞是实现系统进化升级的跃迁路径。如同人类思维产生的过程，一个人脑中的1 000亿个神经元发射掀起了一阵电力传输，其频率可以达到每秒100次。海量的有形神经元细胞捕捉电子进行传导，却形成了无形的、不可捉摸的想法和思维。激活数据学中，高度活跃、高能值的数据单元相互融合聚变产生新的更高阶的数据能量带动系统的整体跃迁，并最大限度地释放数据价值。激活数据学中的各运行规律在块数据系统中相互作用，不断循环往复，这一过程中伴随数据价值的放大和再造，持续推动整个体系螺旋式地进化上升。

人工智能的发展是激活数据学的技术基础，智能碰撞就是块数据与人工智能共舞。随着“无处不在的计算”时代的到来，我们已经进入了一个崭新的智能机器时代。在未来几年内，人工智能和机器人给世界带来的影响将远远超过个人计算和互联网在过去30年间已经对世界的改变。随着人工智能的发展，也许在不久的将来，参与智能碰撞的主体将不只局限于人类。人机智能交互将成为智能碰撞的一种重要互动模式。在这个变化的过程中，首先需要调整的是人类与机器（人工智能）之间的相互地位，转变传统的“人控制机器”的思维定式，以平等的视野去挖掘双方的最大潜能；其次是人类和机器之间的合作关系，不再像过去那样将重心放在人机间的相互博弈上，这将会是人类认知和科学发展的历史性突破；最后是优势互补，打破人脑传统获取知识信息的能力局限和机器的智能盲区，形成平衡所有理智和情感因素后的最佳结论和判断，放大数据的价值。

第三节 基于复杂理论的激活数据学模型探索

（一）块数据与适应性主体

块数据，由个人数据、法人数据、社会数据及国家数据这4类数据构成，是个人、法人、社会和国家4类主体在虚拟世界的数据映射。复杂理论把“适应性”这个生物学概念引用到系统科学的研究领域。由于个人及法人都是具有智能性、自制性及社会性等适应能力的“活”个体，且社会和国家都是在无数个这样的“活”个体相互适应中形成的，所以，激活数据学把块数据抽象为具有适应性的主体是可行的。

块数据作为一个复杂适应系统，具有适应性主体的7种基本现象，

具体表现为聚集、非线性、流和多样性这4个“常见”适应属性，以及标志、内部模型和积木这3个与适应相关的基本功能结构（基本机制）。同时，块数据也具有区别于一般复杂适应系统的特性，主要表现为块数据的聚集是双向且可逆的；块数据的流动具有非耗散性；块数据的内部模型是相对稳定的动态结构，不会凝固。

表5–1 适应性主体与块数据特征对照

现象	适应性主体	块数据	共同点	差异性
聚集	（1）简化复杂系统的一种标准方法，聚集相似的主体成类；（2）简单主体的聚集相互作用，会涌现出复杂的大尺度行为	（1）数据关联融合的基本方式；（2）数据与功能结构以及功能结构之间进行重构的主要手段	构建复杂适应系统的主要手段和基本特征	适应性主体的聚集是单向不可逆的，而块数据的聚集是双向且可逆的
非线性	主体具有主动性和适应能力，它们的属性在发生变化时，并非遵从简单的线性关系	运行结果具有非线性特点，即输出的不可预测性	产生复杂性的内在根源	是适应性主体发生变化的特性，是块数据的基本属性
流	众多主体相互作用过程中的某种资源的流动	数据资源的流动及数据流	流动因时而异，随着主体的适应和不适应而出现或消失	块数据的流动具有非耗散性，即可逆且不一定伴随资源的交换
多样性	种类多样	动态模式	不断适应的结果	块数据自组织的必要条件，较一般适应性主体更宏观
标志	方便对主体及机制的识别、筛选及相互适应	数据搜索、关联融合、自激活、资源配置及智能碰撞的基础条件	贯穿始终的重要机制	
内部模型	当适应性主体接收到大量输入时，会选择相应的模式响应，被选模式最终会凝固成某种特定的功能结构	内部功能结构	实现某项功能的机制	块数据的内部模型是相对稳定的动态结构，不会凝固
积木	通过改变内部模型的组合方式而形成	高层次内部功能结构	原有积木的重新组合增加了系统的复杂性	对于块数据，积木与内部模型之间没有明显的界线

（二）信号、机制和参数

信号、机制和参数是激活数据学研究的三大要素。

第一类要素，信号。信号是能够在块数据系统的功能结构中扩散并引起系统状态变化的“消息”。信号的基本来源包括外界的刺激（External Stimulus，ES）、内部的运行（Inner-Operation，IO）和不同层级的互反馈（Interactive Feedback，IF）。在基于复杂理论的激活数据学模型内，信号是一个不断变化的量，其类型随系统特性的变化而变化，但是其共同职责是不变的，都是对“块内”的数据和内部机制发出指令。对数据发出的指令主要包括识别标志、相互适应、相互聚合以及涌现功能机制；对内部机制的指令主要包括控制现有机制进行拆分、关联及融合，以涌现出更高层次的功能机制，甚至引发系统整体状态的切换。由此，信号被定义为“触发”块数据系统进行功能及状态切换的“信息”。由于信号（Signal，S）的表现形式是数据引力波（Data gravitational wave, Dgw），那么，对于信号的形式化描述为

$$S = Dgw::\{ES, IO, IF\}$$

其含义为，信号是以数据引力波的形式对系统产生作用的，且信号具有感受外界刺激（ES），指挥内部运行（IO）以及接收各种反馈（IF）的能力。

第二类要素，机制。对应激活数据学表现出的运行规律，激活数据学模型构建了5种运行机制，分别是刺激响应机制、过滤筛选机制、状态转换机制、资源分配机制以及进化机制。其中，刺激响应机制是块数据的基本功能，过滤筛选机制是激活前进行预处理的功能结构，状态转换机制是块数据在潜动源、动源和热动源三种状态间转换的功能结构，而资源分配机制和进化机制是在块数据系统的高度灵敏性被激活后才能具有的高级功能结构，主要负责优化资源配置及数据价值的创造与放大。

其一，刺激响应机制。刺激响应机制（Stimulus Response, SR）对应的激活数据学运行规律包括数据搜索、关联融合和数据自激活，出于赋予块数据以能动性的目的，刺激响应机制具有三种功能：刺激接收功能、刺激处理功能和刺激效应功能。刺激接收功能负责接收刺激输入的信号，刺激处理功能负责对接收到的信号做出反应，而刺激效应功能则负责按照信号内容对外界环境产生相应的作用。三种功能协同的过程可以被形式化描述为：

$$输入：X = \{x_i\}，i = 1，2，3，\ldots n;$$

输出：$Y=\{y_i\}$，i＝1，2，3，…n;

$$Y=F(X)$$

其中，F表示刺激（X）到反应（Y）的映射，即接收输入的刺激自变量，处理刺激，输出反应因变量。

其二，过滤筛选机制。过滤筛选机制（Filtering Selection，FS），对应的运行机制包括数据搜索和关联融合。在块数据自适应过程中，“块内”数据会基于不同目的对其他数据进行过滤，并筛选出适应性最强的数据进行关联融合，以生成内部功能结构。在对内部功能结构完成应用后，相关数据会根据应用的结果，积累筛选数据及关联融合的经验，并根据已有经验对已经与其产生关联的数据进行拆分、过滤、调整或替换。可以说，这是一个“块内”数据通过定量积累经验以完成定性目标筛选的过程，且整个过程可以应用数学中的组合公式进行形式化描述，即输出候选数据列表：

$$C(n, m)=P(n, m)/m!$$

其含义是，在包含n个候选数据的集合中，取m个数据进行关联融合的全部取法，即形成$C(n, m)$种数据搜索及关联融合的可选方案。

其三，状态转换机制。状态转换机制（State Transition，ST）对应的运行规律是数据自激活，是激活块数据在潜动源、动源和热动源三种状态间转换的功能结构。基于复杂理论，对状态转换机制运行过程的微观描述是，驱动块数据系统从亚临界状态进入临界状态，再到连接着临界状态与超临界状态的引爆点的过程。到达引爆点的块数据系统的灵敏程度最高。激活的过程实质上是块数据从旧状态向新状态转换的过程，所以，可以把状态转换机制假设为一个转换函数，即f（ ）；用集合S表示旧状态的集合S＝{潜动源，动源，热动源}，即函数的自变量；用S'表示输出的热动源状态，即函数的因变量。于是，状态转换函数可以定义为$f(S)$，其形式化描述为

$$f(S)=S'$$

其中，S'代表只包含热动源一个元素的集合，因为函数的特性是，无论输入端输入的状态是什么，其输出状态都一定是热动源状态，即块数据系统被激活。

其四，资源分配机制。资源分配机制（Resource Allocation，RA）对应的激活数据学运行规律是热点减量化，是块数据对有限资源进行最优配置的功能。其过程可以被简述为，通过识别输入的热点数据的标志，按照系统的需求，对热点数据资源进行合理化配置的过程。那么，其形式化描述为

RA∷{RID, Ri, Rf, RS}

其中：*RA*表示资源分配机制；*RID*表示热点数据的标志；*Ri*表示输入的热点数据的集合；*Rf*表示对热点数据的配置函数；*RS*表示资源分配机制的状态，也是资源分配机制向块数据系统进行反馈的关键变量。

其五，进化机制。进化机制（Evolutionary Mechanism，EM）对应的运行规律是智能碰撞，期望实现的是被激活的数据单元相互融合聚变产生新的更高阶的数据能量带动系统的整体跃迁，并最大限度地释放数据价值。进化机制是以数据价值最大化为根本目标，通过变异和突变的方法促使块数据系统完成自身功能进化的过程。进化机制的最主要步骤，是根据功能结构的使用结果，给每一个完成应用的功能结构贴上“适应指数标签”，再对带有标签的功能结构进行拆分及重组，以实现进化块数据系统整体功能结构的目的。所以，对进化机制基于“遗传算法”的描述为“从经验中积累，从已有规则出发，通过选择、交换、突变等手段使块数据的功能得到充分的进化。”注

进化机制可以被描述为“块数据的遗传算法”。完成进化可以分为以下5步。第一步，聚集。产生初始的功能结构的集合，即拆分多个已完成使用的功能结构。第二步，评估。对功能结构的适应性进行评估及预测。第三步，选择。在被拆分的功能结构集合中选择优良的结构片段。第四步，交换。用已选结构片段重组更优秀的功能结构。第五步，变异。随机选择现有功能结构，改变其中的某一片段。这种变异操作具有一定的概率成为突变，形成具有奇异功能的功能结构，可以作为实现智能碰撞的一条可行途径。

总之，进化机制就是在给定初始功能结构的前提下，以“迭代”注的方法实现对已有功能结构的进化，且进化是一个非线性的过程。那么，对于进化机制的形式化描述是

EM∷{P（J），N，F，S，G，T}

其中，$P(J)=\{J_1, J_2, J_3, \ldots J_n\}$，表示初始机制集合；$N$表示集合的规模；$F$表示适应函数；$S$表示选择算子；$G$表示遗传算子，包括交换算子（$OC$）、突变算子（$OM$）；$T$表示阶段性的终止规则，进化机制阶段性终止工作的时间是系统达到短暂的有序与无序平衡状态的时刻T。

第三类要素，参数。由于从空间来看，自组织临界相当于一个亚稳定的有限维动力学吸引子，并且自临界组织不需要通过调节参数来逼近临界点。所以，在基于复杂理论的激活数据学模型中，参数是每个系统的特定属性，对系统本身而言参数是常量。对于多个系统而言，参数则是变量，是决定不同系统内部信号和机制差异的锚定因子。另外，块数据系统由不定数量的参数共同作用，其功能属性各异，所以无法对参数进行特定的形式化描述。只能把参数定义为块数据内一种决定信号和机制差异的功能常量。

（三）激活数据学模型的形式化描述

激活数据学模型的三大基本要素包括信号、机制和参数。信号，是能够通过数据引力波在“块”内扩散并引起系统状态变化的“触发消息”。机制，是具有复杂性的动态功能结构，为了对应激活数据学的5种运行规律，激活数据学模型描述了5种主要运行机制，包括刺激响应机制、过滤筛选机制、状态转换机制、资源分配机制以及进化机制。参数，是系统特定的差异性属性，是决定信号和机制的锚定因子。

块数据具有在非平衡和非线性共同作用下的高度灵敏性，并在宏观层面表现出长程的秩序并演化出多样化的自激活、自流程、自组织状态。激活块数据的过程就是使块数据以自组织的形式不断趋向引爆点的过程。整个运行过程是在信号、机制及参数三大基础要素的协同作用下完成的。其形式化描述为

$$\mathrm{Active}::\{\mathrm{AS}, \mathrm{AM}, \mathrm{A\Phi}\}$$

其中，$Active$表示激活过程的形式化描述，AS为激活信号，AM为激活机制，$A\Phi$为激活参数。

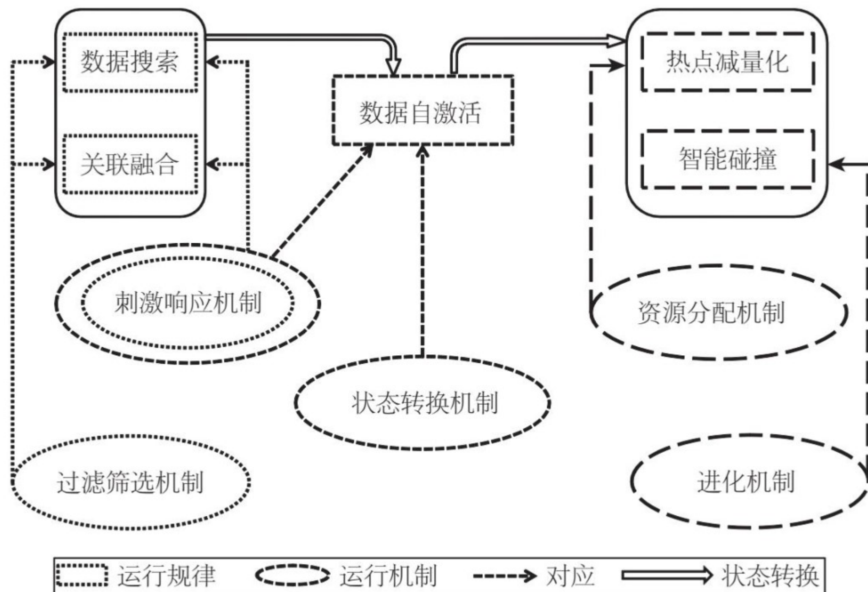


图5–2 激活数据学模型机制作用结构图

从激活数据学模型的基本构成出发，可以将块数据具有的基本特征概括如下。第一，块数据各机制之间具有广泛的联系，构成一个紧密体系。单一机制的变化都会受到其他机制变化的影响，并会引起其他机制的变化。第二，块数据具有多层次、多功能结构，每一层次的机制均成为积木构筑其上一层次的机制，同时也有助于系统某一功能的实现。第三，块数据是开放的，它能与环境中的信号紧密联系并相互作用，不断适应环境。第四，块数据是动态的，它处于不断发展进化之中，而且对于未来的变化具有预测能力，并迎合未来的变化。

综上所述，激活数据学以发现块数据内海量复杂数据的潜在关联和预测未来为目标，以复杂理论的系统思想为主要范式，不仅论证了块数据即是适应性主体的探索前提，而且根据激活数据学的运行规律，发现了引爆块数据的基本流程：通过数据引力波对信号进行传导，以控制多种自激活机制共同作用，来实现驱动块数据自流程化运行，在非线性与非平衡的共同作用下，自组织成为具有高度灵敏性的复杂自适应系统。

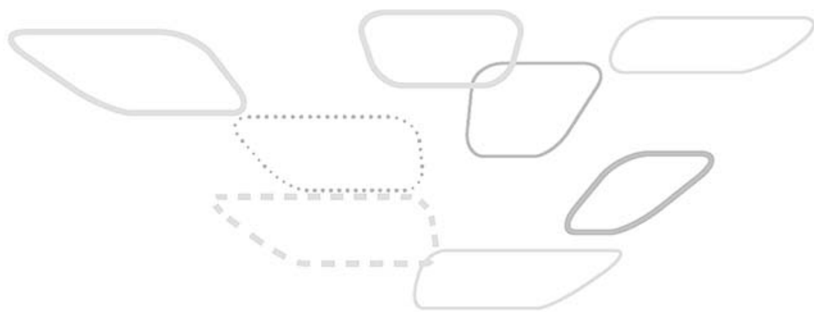
---

1. 张小彦：《数据融合：大数据分析的瓶颈》，财新网，2016年3月。

2. 许志国：《系统科学与工程》，上海：上海科技教育出版社，2001年。

3. 迭代是重复反馈过程的活动，其目的通常是为了逼近所需目标或结果。每一次对过程的重复称为一次“迭代”，而每一次迭代得到的结果会作为下一次迭代的初始值。

第六章
块数据价值链



块数据的真正价值在于应用，块数据在商用、民用、政用等领域都有着广阔的应用空间和前景。孤立的数据本身不存在价值，只有通过数据与人的互动、关联和融合，实现数据与人、物、事之间关系的重构，才能最终形成全新的价值链条——块数据价值链。

块数据价值链的研究重点是基于商业的全产业链、基于社会的全服务链和基于政府的全治理链的形成，通过块数据价值中枢，以数据流引领技术流、物质流、资金流、人才流、服务流，推动产业业态、社会形态、国家治理模式等一系列变革，实现数据价值的最大化。

块数据价值链的本质是实现超越资源禀赋的新的价值整合。通过数据驱动，实现产业重构和价值再造；通过需求识别，实现公共服务需求导向的转变以及模块化供给体系的构建；通过数据感知与预测，实现块数据治理的精细化和精准化。

第一节 块数据价值链

（一）价值与数据价值

价值是无差别的人类劳动。马克思劳动价值理论强调，价值是满足社会某种需要的效用，是形成产品价值的前提，任何产品，如果对社会

没有效用，那么生产它所花费的劳动，就是抽象劳动的浪费，因而不形成价值。

关于价值的本质，还存在多种观点。客观主义者说，价值是我们在事物、人或境遇本身中所发现的某种东西。价值是寓于客体内部的一种性质，它并不存在于观察它的人本身之中。主观主义者说，价值与其说是对象的某种属性，不如说是主体的一种创造。关系主义者说,价值既不是我们的独立创造，也不是行为或对象的独有属性，价值是从人与行为或对象之间的某种关系中产生的，是主体和客体之间相互作用的结果。注

从信息社会到数据时代，数据不仅成为信息的代名词，而且成为比信息更大范畴的概念。1948年，香农从热力学角度观察信息学，让信息的不确定性可以被量化测度，并最终将其定义为“信息熵”。对比“信息熵”的概念，我们认为，数据价值就是减少事物不确定性的量度。

对于这种不确定性的描述，车品觉注将之进一步细分为五大价值，分别是识别与串联（辨别关系和身份），描述（刻画研究对象特征），时间（记录行为），预测（分析趋势），产出数据（数据整合）。

伴随着海量数据的涌现，数据价值还更多地表现出隐匿性和持续性的特征。

隐匿性。数据源更多以混沌的状态作为初始存在，价值本身并不会主动地表露，对数据的粗加工价值有限，需要将这些混沌的数据分解为数据元状态，寻找隐性连接才能发掘更多价值。这需要高精度的复杂运算。

持续性。有形资源和资产由于其物质本源的限定，会因损耗而产生贬值。而数据资源由数据流产生，继承了持续性的特点，即不会因为使用次数或时间推移而发生增减变化。数据还有可复制性和交换性，它能够在多个空间维度中有更多增值。

（二）块数据价值链

传统价值链理论。**1985**年，美国哈佛大学商学院教授迈克尔·波

特在《竞争优势》注一书中首次提出了“价值链”的概念。他认为价值活动贯穿于经济组织行为活动的供需链环节，采购、生产、仓储、销售、售后等活动都能为卖方提供有价值的劳动产品，这些活动价值会以价值链的形成呈现出来。企业在市场活动中通常有两种方法扩大其竞争优势，一是强化其主营业务，增加其价值链中某一环节的议价权；二是兼并上下游，减少其价值链中其他环节的成本消耗。企业经营者在战略层的部署多是从这两种角度切入展开。价值链理论的实际效用早已超过了方法本身，进而影响了众多的研究者和战略决策者。



图6–1 价值链模型

虚拟价值链。1995年，哈佛商学院的杰弗里·F·雷鲍特和约翰·J·斯维奥克拉这两位学者在《开发虚拟价值链》一文中首次提出“虚拟价值链”的概念。信息技术对生产工具和生产方式进行革新，进而改变了其属性。在传统的经济行为中，实业制造是其核心的价值增长点；而在虚拟世界中，企业通过对信息的采集、加工、存储、组织应用也能够产生大量的经济收益。虚拟价值链的概念突破了传统价值链仅涵盖实物的范畴，将信息、知识、数据等虚拟资源列为价值链增值的重要元素。



图6–2 虚拟价值链模型

李长玲等在构造企业信息价值链模型的研究中，提出了综合信息增值手段以及方法，将虚拟价值链具化为理论模型。注



图6–3 信息价值链模型

《知识价值链研究现状分析》一文提出，知识的获取、创新、保护、整合和扩散是知识价值链的核心流程，并由此建立了包括知识管理基础以及认知过程管理的知识价值链模型。注

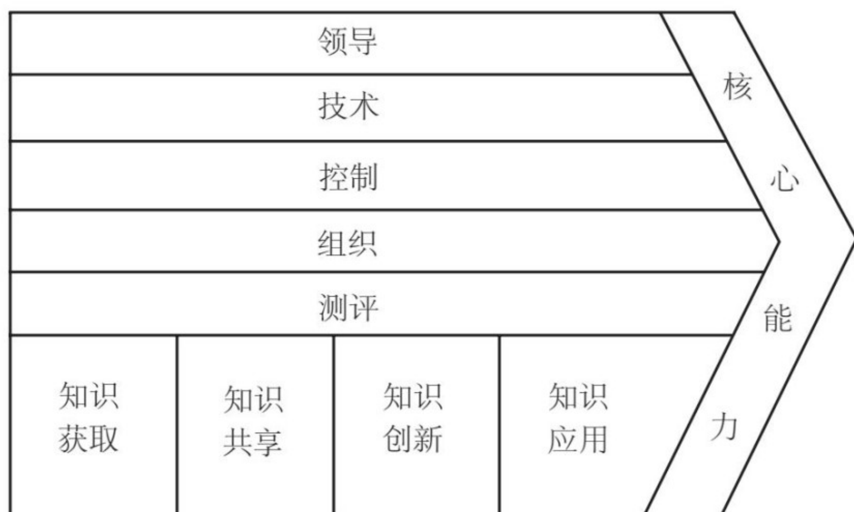


图6–4 知识价值链模型

价值网理论。基于网络平台和网络结构，很多学者对新型的价值链提出了探讨。这种新型价值链不同于以往的由增加价值的成员构成的链条，它是一种由虚拟企业构成的网络，被称为价值网。价值网的概念突破了原有价值链的范畴，它改变了价值链单向线性的组合传递结构。

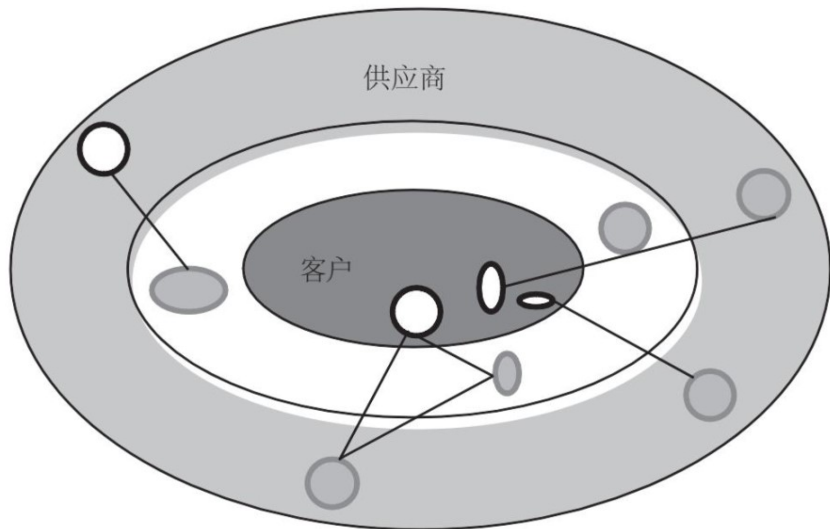


图6–5 价值网结构模型

全球价值链理论。全球价值链理论是伴随着全球经济一体化应运而生的，主要研究跨国企业在全球范围内开展生产经营活动和利益分配的影响因素与内在规律。全球价值链包括4个维度：投入——产出结构、空间布局、治理结构和体制框架。全球价值链理论描述了价值链的空间分离和全球空间再配置之间的关系，拓展了价值链分析的空间视野。

块数据价值链理论。伴随着数据经济、网络社会、移动互联、人工智能等诸多世界经济发展动态和趋势，价值链理论的重心不断地转移，呈现出从实物到虚拟、从线性到非线性、从单个组织内部到无限边界等诸多变化。块数据价值链理论的提出正是这一理论深化的最新成果。

块数据价值链的任何环节创造价值都要涉及5个步骤：采集、传输、存储、分析和应用数据。价值创造主要由基本价值活动（在生产或服务的各个节点都会产生数据，串联5个环节汇总形成数据流）和增值性活动（以数据流引领技术流、物质流、资金流、人才流、服务流，优化资源配置；在各节点的价值增值过程又会产生更多的新数据）两部分来完成，而这些相互关联的价值活动便构成了价值链。在价值链的价值增值过程中，产生产业业态、社会形态、国家治理模式等一系列的变革，通过全新的更大范围的价值体系重构，实现新的数据价值的最大化。

块数据价值链具有如下三个重要特征。第一，新的价值载体。数据流成为与资金流、物质流和人才流同等重要的价值载体。大数据作为一种新的基础性战略资源，可以提高企业和公共部门的生产力和竞争力，以及提高消费者福利。第二，新的传递机制。

从价值传递形式看，链式结构被拓展成网状结构，形成价值网络。第三，新的配置范围。从资源配置的范围看，由于块数据的泛边界性，使得突破原有特定组织和地域限制，在更加广阔的范围内调动资源实现优化配置成为可能。

（三）块数据价值链模型

块数据价值链模型强调，通过在各节点上采集、传输、存储、分析和应用数据，在数据的价值创造与价值传递过程中实现价值增值；数据流经过块数据价值中枢的价值发现与再造，产生的数据驱动力，带动和影响技术流、物质流、资金流、人才流、服务流的流向，实现资源优化配置，最终催生了包含基于商业的全产业链、基于社会的全服务链和基于政府的全治理链的多元价值体系。

块数据价值链理论将决定价值的几个基本要素的运动称为数据价值流，由6个方面构成，即数据流、技术流、物质流、资金流、人才流、服务流。其中，数据流是核心资源，决定和影响其他要素流的流向和组合关系。这种运行方式突破了传统价值链的链条式结构和线性思维，各结点间相互衔接、融合以及动态互动，共同构成了一个围绕块数据价值中枢的网状结构。



图6–6 块数据价值链模型

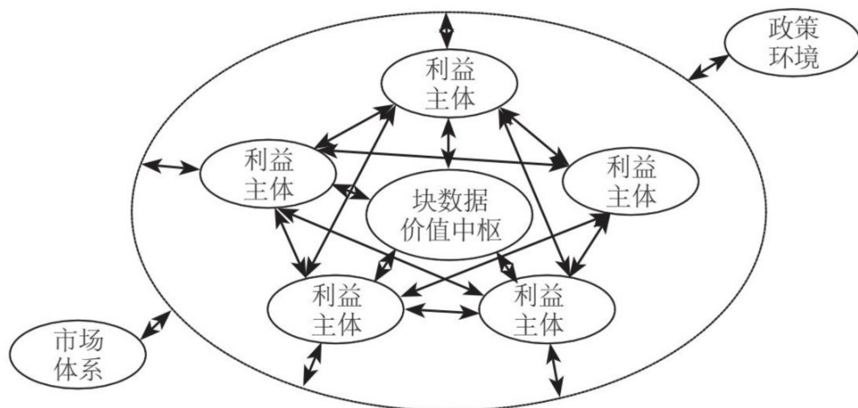


图6–7 块数据价值流的网状结构

块数据价值链的结构特征可以概括为：

块数据是价值中枢。块数据通过对交易市场和政策环境下的用户行为和需求进行深度挖掘，组织加工后反馈给供给侧和需求侧，这个过程越是高度灵敏精确，越能加快市场主体对经济形势变化的响应速度。在这个价值链网络中，通过数字感知、传输、分析和处理技术，识别用户需求以及市场和政策环境的变化，是激活价值链网络的关键。

用户价值是核心。块数据价值链的最大价值是应用，是解决现实世界中的各种难题。从商业、社会到政府，都可以找出基于块数据价值链的解决方案。无论是基于商业的全产业链、基于社会的全服务链还是基于政府的全治理链，都把客户看作价值的共同创造者，即用户是价值流动的起点，同时通过对各环节的渗透融入价值体系之中，并把用户的要求作为价值取得的最终决定因素。

关系网络是支撑体系。在这个网络体系中，数字化的关系是协调、配置、影响各相关主体行为的关键，使得分布于网络上的各节点能够迅速组织进而快速响应用户需求。

生产和服务提供商是微观基础。在市场主体的微观活动中，更强调价值网络成员之间的关联协作，充分发挥彼此间的正外部性，提高资源配置的作用余地。当生产商和服务提供商将资源集中于主营业务，会形成更强的话语权，在价值网络中的融合度和效度也会越高。

第二节 块数据全产业链

（一）数据驱动和产业重构

资源学派认为，资源要素是一个组织从事任何生产、研发或者服务、管理所必备的条件，是产业价值链竞争优势的重要来源。能给组织带来长期可持续竞争优势的资源需要具有战略性，即具有稀缺性，一般性资源只是一个组织存续的基础，它不能保证组织能因此获得超额利润。注在不同的经济发展阶段，组织的战略性资源是不同的。土地在农业文明时代是战略性资源，到了工业文明时代，土地不可能给组织带来竞争优势，土地便从战略性资源变成了一般性资源。而在数据经济时代，数据成为最重要的战略性资源，深刻改变了传统的产业形态和管理模式，进而影响了整个人类的价值体系、知识体系和生活方式。

大数据作为基础性战略资源，将助力产业转型升级。中国经济发展新常态的重要特征之一是，“模仿型排浪式消费阶段基本结束，个性化、多样化消费渐成主流”注。从全球范围来看，德国工业4.0、美国先进制造伙伴计划、中国制造2025，都是在不同程度上响应新变化的前沿战略，并且给出了明确的目标愿景和路线图。它们共同的特点是：都以数据为核心驱动力。数据把价值链的各个环节连接在一起，首先是通过技术创新，把用户需求转化为数据模型；其次是价值链相关方之间的网络关系数据化，实现研发、生产、销售、运输、售后等环节的数据协同。可以说，数据驱动是产业转型升级的强力引擎。

无论是处在转型升级变革阶段的传统行业，还是基于新一代信息技术应用构建的新兴行业，都必须从顶层设计开始，构建“数据驱动战略”。数据驱动战略的关键是“让数据流动起来”，通过数据流带动资金流、物资流、人才流等要素资源在价值链中快速流动，实现价值的创造与增值。例如，在阿里巴巴主导的产业链中，从产品代理、店铺融资、加工制造、定价、推广、销售和售后服务，都能够获得提取自用户的行为数据，为店主提供免费或增值的数据分析服务，提高销售主体的经营水平，进而产生更多的用户数据，这些新产生的数据又能在平台上产生更多的项目和价值。产业的数据驱动如同巨大的飞轮，初始启动较为艰难，需要持续不断地施加推动力，不断克服系统阻力，循环从初始的缓慢状态变得越来越快，在飞轮快速旋转之时，产生惯性作用，此时只需要少量的推动，就会产生巨大效果。数据的增量规模和流动速度与数据

流的价值创造相互促进，形成正向反馈的循环。

（二）块数据全产业链体系

基于块数据的驱动，各产业部门之间依据内在的关联规则，形成新的关联关系，从而对传统产业结构和产业体系进行颠覆性解构和重构，形成新的全产业链。块数据全产业链是包含了数据核心层、数据产业关联层和数据产业衍生层三个相互联系、相互依存的子系统的复杂价值系统。其中，核心价值活动由数据生产、数据采集、数据存储、数据分析、数据交易和数据应用这6个核心环节构成；关联层由核心层各环节的上下游紧密关联产业构成，呈现出智能终端产品、电子材料和元器件、集成电路、呼叫中心等业态；衍生层是核心层和关联层相互交融后在其他领域的衍生应用形态。

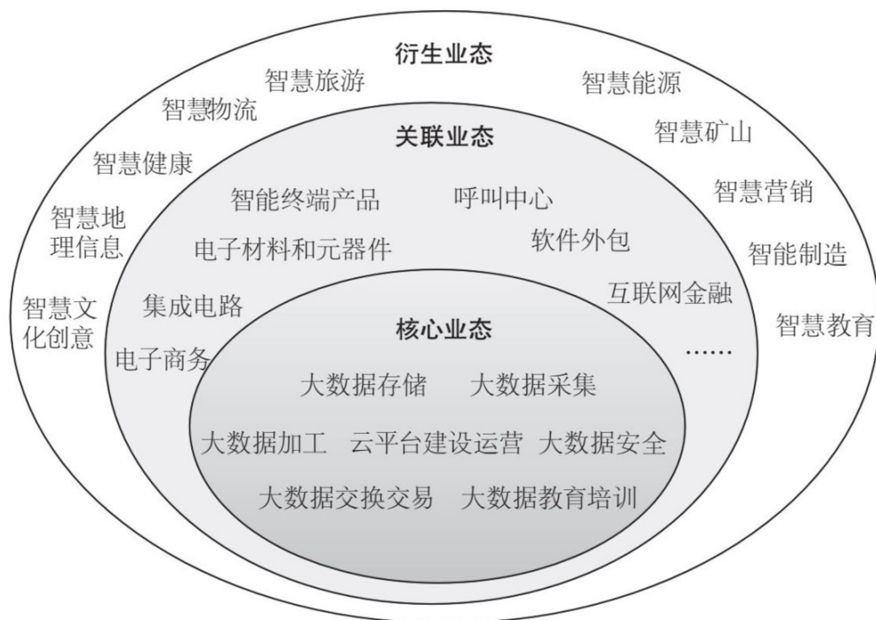


图6–8 块数据全产业链结构图

在块数据全产业链中，呈现传统线性产业链向网状产业链转变的趋势，即由“价值链”到“价值网”。价值网络的形成过程，是由线性价值链的相交形成结点，结点数量增多，最终形成网络。价值链是主体间价值流动的路径，任意可行的连接都是一条价值链，以指数级的增长方式极大地丰富了价值创造的实现方案。相较于单独的价值链，价值网的价值放大效应更加明显，可以获取的总体竞争优势超过了单个组织能力的总和。块数据全产业链的网络化结构实质上是对产业链观念的变革，即由

产品供应链转向需求链，由产品驱动转向需求驱动，价值链由前端推动转向终端驱动。

（三）块数据产业的价值再造

从价值链理论出发，数据经济时代的产业转型升级，关键在于实现从价值链低端向中高端的转移，提高产业发展的质量、效益和竞争力，在于创新要素的集聚和结构的优化。通过以数据为核心的创新要素的进一步解放、流动和优化配置，实现价值链再造和升级，最终提高全要素生产率。

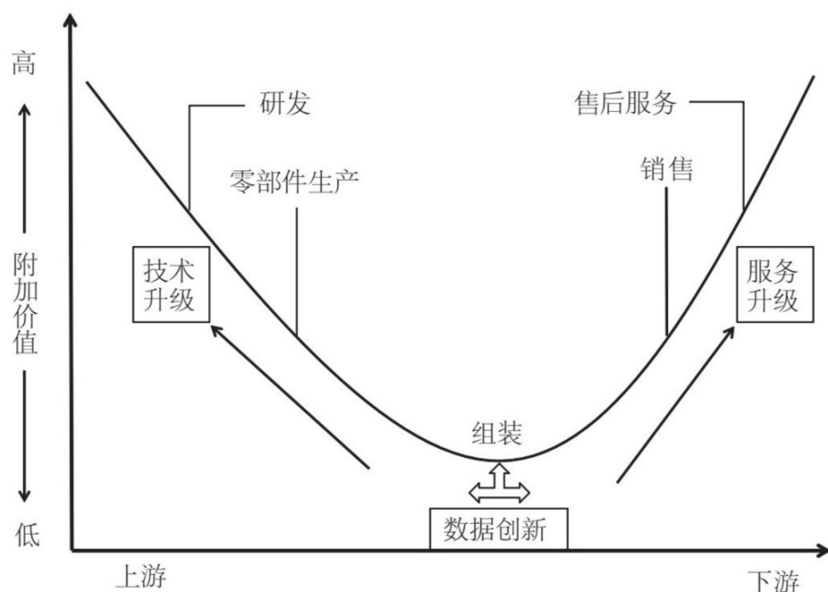


图6–9 产业链升级的“微笑曲线”

全产业链升级与传统产业改造。块数据全产业链的层级覆盖范围，不仅仅是指数据处理、云计算、物联网等新兴产业和业态，还涉及制造业、服务业等众多传统领域。通过块数据的乘数效应，整合创新要素，向微笑曲线的两侧延伸，促进价值链升级。

数据创新体现在产业链的各个环节，例如，海尔率先实施工业4.0战略，开展个性化家电定制服务，并根据网络反馈强化新产品的研发导向；而淘宝店客服会根据不同的用户信息对新客户、回头客和固定客源提供不同的营销服务。事实上，有关“从劳动密集型产业到资本、技术密集型产业，尤其是战略型新型产业，是产业转型升级唯一方向”的论调，

在实践中是经不起推敲的。迈克尔·波特认为，任何传统产业只要加上知识和技术，都可以成为一个具有国际竞争力的技术密集型产业，这是关于传统产业升级最到位的说法。

在微笑曲线中，越偏向个性化、创造性的环节，其价值空间的增量就越大。从研发设计、生产制造、物流仓储到售后服务等产业链组成的产业中，知识需求主导的附加价值变化幅度也会放大。尚品宅配是一家家具生产企业，主要为用户提供家具的个性化解决方案。通过对数据的运用，尚品宅配实现了传统产业价值链的再造。一方面，是对客户个性需求数据化。它拥有一个庞大的数据库，包含全国重点城市约2 000家楼盘的10万个房型，以及相关的居住空间解决方案；设计信息库中不同款式、结构、材质的多系列产品过万种，两者结合形成高效丰富的设计能力。另一方面，全国的订单汇总之后，开始快速地“拆单”和“并单”，每件家具大概被分拆为几百个零部件，在48小时内完成从汇集订单、识别加工图、再到封边、钻孔等一系列操作，通过数据管理中心和柔性生产体系，可以让尚品宅配的差错率从传统厂商的30%下降到3%，生产效率是传统厂商的10～20倍。

块数据全产业链升级的跨越式路径。美国杜克大学的杰罗菲最早从事价值链升级理论的研究，提出了转型升级可分为4个层次：包括产品层次升级，经济活动层次升级，部门内层次升级，部门间层次升级。2000年，美国汉弗莱和施迈斯提出以企业为中心的分类法，即分为工艺流程升级、产品升级、产业功能升级和价值链升级。注



图6–10 传统产业价值链升级的层次划分

传统价值链理论刻画出了一条由“工艺升级到产品升级，再到产品功能升级进而价值链升级”的渐进式路径。最初的切入点是将生产流程的工

艺工序切割细分；然后进行技术积累，当积累到达一定水平后，利用主营业务的基础平台做横向产品研发延伸，增加其内部价值，兼并上下游的产业环节减少成本；最后，让渡一部分被压缩的低利润环节，将企业的价值核心凝结在数据创新和核心能力上，最终实现向更高层级价值链的升级。

而现实的产业升级是一个复杂的系统工程，受到多变的内外部环境影响，在特定背景下可以呈现出跨越式发展的特征。以块数据为核心资源的全产业链，提高了价值链各环节的效率，突破了常规演进规律，实现了跳跃发展。后发地区和部门可以快速融入块数据全产业链，充分利用要素数据化和经济全球化的有利条件，在尽可能短的时间内走上跨越式升级的发展道路。

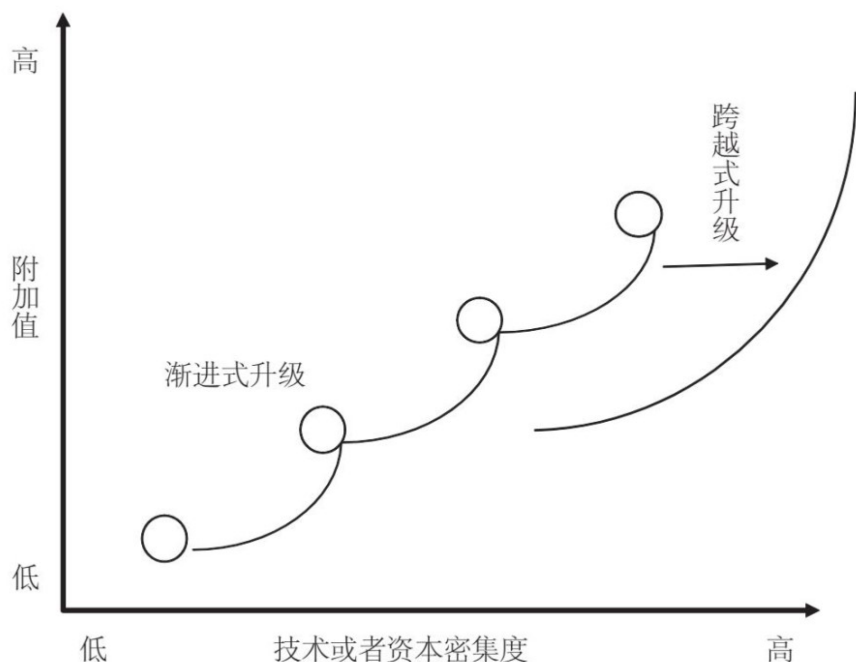


图6–11 价值链渐进式与跨越式升级的路径

第三节 块数据全服务链

（一）顾客导向和需求识别

在行政管理领域，对政府的角色定位有诸多描述，其中，戴维·奥斯

本同特德·盖布勒合著的《改革政府：企业家精神如何改革着公共部门》提出一种“顾客驱使”的观点，即将政府的驱动者理解为顾客，而非官僚体系自身。基于这种观点，人们将政府同公民、法人、社会组织的关系转换为供需关系，即政府应当以顾客需求作为第一价值驱动。从古至今，政府与公众的关系一直是治理与被治理的关系。而进入网络社会和数据时代，政府的主要职能由管治向服务转移，公共事务的数据化处理能力将受到更为严苛的考量。“公共性”是政府治理的本质特征。政府的“公共性”体现在，其服务对象不是政府机关自身或者社会上的少数人与组织，而是社会上的各类人群与组织。在公共性的主导下，政府必须将满足公众需求作为行为导向，如同企业的出发点是为用户服务一样，公众需求是政府公共服务存在的前提。

对公众需求的分类分析，需要引述日本学者狩野纪昭（Noriaki Kano）关于KANO模型的观点。他将用户需求按照基本型、期望型和兴奋型做了三个层次的划分，这是对公众需求的精确分类。

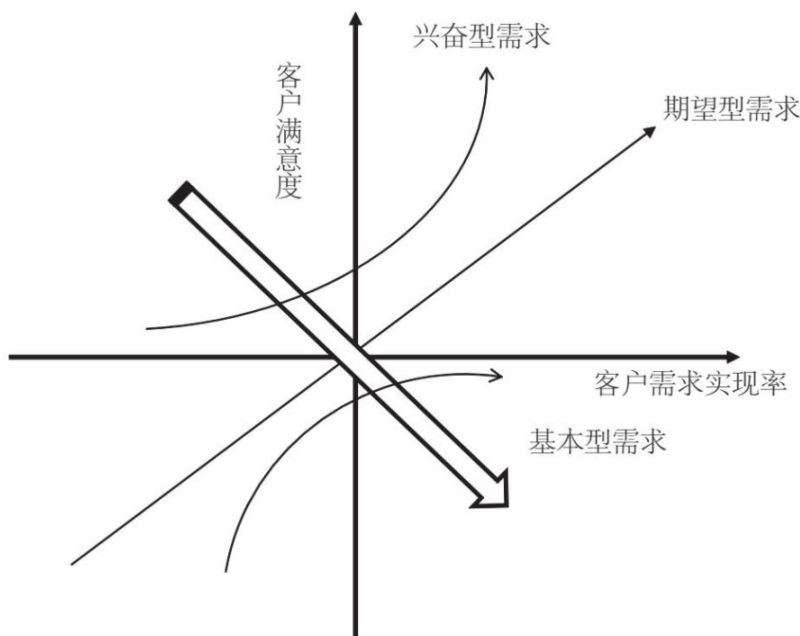


图6–12 KANO模型关于三个层次的顾客需求的定义

第一类，基本型需求，是指产品和服务应具备不可或缺的功能属性，当这些功能属性无法达到使用要求时，就会阻碍用户购买产品或服务。第二类，期望型需求，是指产品和服务的质量会与顾客满意度发生更大关联，通常体现在产品具有区别于同类产品的功能和特性。第三

类，兴奋型需求，是指产品会带来更多的附加价值，这类产品或服务通常提供了超出用户期望值的价值，刺激更多的购买欲和好评。而且，随着时间的推移，兴奋型需求、期望型需求、基本型需求会呈现出向下递进转化的发展趋势。

识别需求是公共服务持续创新的动力。公众需求具有动态性和复杂性，要提高政府公共服务的供给水平，必须明确公众现实和潜在的需要，通过技术和管理的创新来追踪公众需求，把公众“是不是需要、是不是支持、是不是满意”作为衡量政府公共服务供给的标准，从而形成由公众需求为导向的公共服务持续创新机制。

（二）块数据全服务链体系

信息技术改变了人的存在方式，它的变革将改变人类的学习方式、工作方式、休闲方式和生活方式。在这样急剧的技术背景和时代背景的作用下，公众对于公共服务的需求也从现实社会延伸到虚拟社会，多元行为主体对公共服务的内容和实践提出了更高的要求，新的服务理念和服务模式呼之欲出。

全服务链是块数据价值链模型在公共服务领域的具体应用，它围绕块数据价值中枢，从包括公共事业、市政管理、城乡环境、农村生活、健康医疗、减灾救灾、社会救助、养老服务、劳动就业、社会保障、文化教育、交通旅游、质量安全、消费维权、社区服务等领域中，精准识别公众需求，寻找民生痛点，进而构建整合政府、企业、社会组织等多元主体的模块化供给体系，提供层次性多样化的公共服务。

参照KANO模型的分类，在块数据全服务链中，公众对公共服务的需求被划分为基本需求、拓展需求和潜在需求这三类：基本需求是指需要政府对公民和社会组织提供基础的服务内容和功能。这类需求得到满足，满意程度不会提升，而一旦不能被满足，就会引起公众的不满。拓展需求是指政府需要依照公众期望做出改良。这类需求与公众的满意度正相关，需求得到满足，满意度会同步增长，反之则会下降。潜在需求是指公众还没有关注到的需求，如果公共服务供给部门能够前瞻性地预测出公众的此类需求并提前供给，将使公众的满意度大幅提升。在此三类之外，如果公众对政府是否提供某项服务的内容和治理漠不关心，就把这类服务定义为“无关特性”。不过，由于公众需求的差异性，某个群体漠不关心的项目，另外一个群体却可能非常关注，出于对群体差异性

和动态性的考虑，无关特性也被列为需求分析时需要考虑的一个类别。

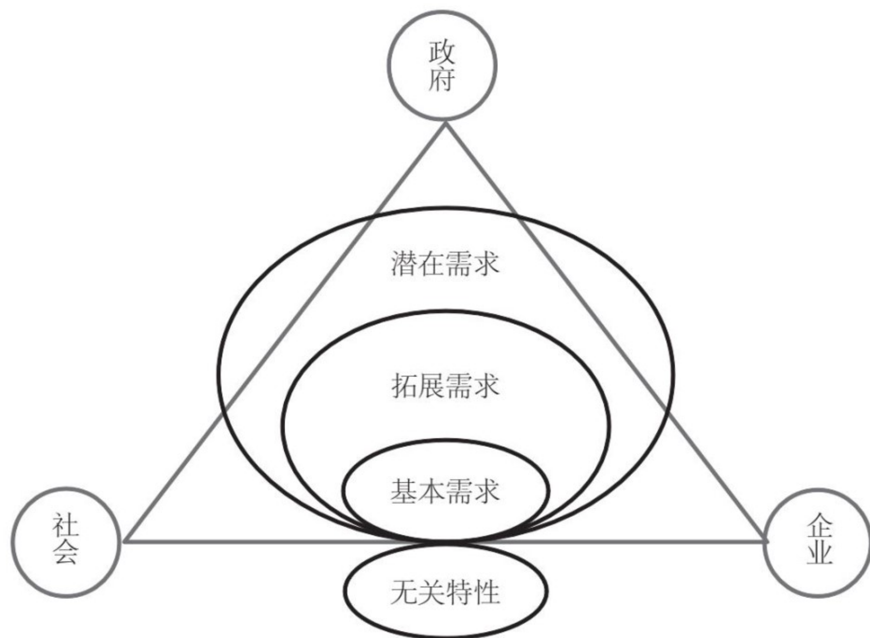


图6–13 块数据全服务链结构图

以上4个类别构成了政府公共服务需求分类的体系结构，其中基本需求、拓展需求和潜在需求这三类公众需求存在一种双向层次递进关系。一方面，当公众对需求的要求不断提高，在较低层次需求的满足后会转向对较高层次需求的追求；另一方面，服务满意度的效应会不断降低，较高层次的需求得到满足后就会逐渐转化为较低层次的需求。另外，随着时间的推移，无关特性中的一部分也可能转化为潜在需求。

在政府公共服务的供给实践中，需要构建多元主体的综合供给体系，针对公众需求提供层次性的供给方式。首先，务求圆满覆盖基本需求所涉及的内容和领域；其次，对拓展需求进行重点攻克；而潜在的个性化需求，在综合考虑成本和收益的情况下，积极引导社会和市场力量介入，实现公共服务供给的有效补充。针对无关特性的需求，要衡量其利弊，根据其影响的深度和广度来最后确定。

（三）块数据服务的模块化供给

公共服务供给与多元的公众需求之间的矛盾，曾经主要在“市场和政府”的关系框架内寻求应对，并由相互对立走向融合。随着相关研究与实践的展开，进一步打破了“市场——政府二分法”，引入政府、第三部

门、私人组织等多元主体，通过模块化供给机制提高公共服务供给能力的思路渐成共识。块数据全服务链正是基于这一思路的最新解决方案。在全服务链中，块数据价值中枢作为多级子模块之一，需要与其他模块实现协同，共同对多元化公共服务需求做出敏捷的响应，而模块化生产方式利用模块分解实现模块规模经济，再通过模块组合实现供给弹性来提高供给能力和敏捷程度。从具体构成看，块数据服务的模块化供给过程可以归纳为“一平台，两机制”：

公共服务协调平台。受各服务链环节间彼此独立、信息不对称、地域阻隔等诸多因素影响，在复杂环境中响应的速度和针对性会大打折扣，以块数据为基础的公共服务协调平台就能很好解决这些问题。其主要功能包括：一是信息交互功能，通过供需双方的信息交互消除信息不对称；二是组织协调功能，通过公共服务模块化生产系统，对涉及供给、生产、存储、流通等相关各方提供科学指导，让各模块以协调融合的方式运行；三是反馈优化功能，通过采集供给产品的反馈意见，评价供给体系与需求变化之间的差距，进而优化调整公共服务供给。

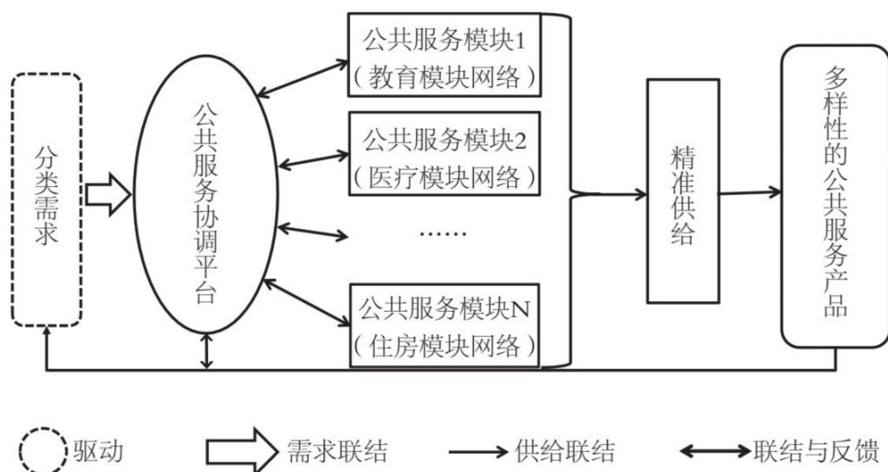


图6–14 公共服务模块化供给的结构流程

分层组合机制。全服务链中的服务和产品供给单元，由基础模块和基于基础模块之上的模块组合共同构成。基础模块以专业化为导向，以发挥规模经济效应为宗旨；模块组合以提高供给弹性为目的，提供多样化且敏捷化的公共服务产品。通过全服务链各不同模块间的分解和组合，公共服务生产与供给体系能够实现较高的供给能力和响应速度。

精准供给机制。与从供给端出发的“推动式”驱动机制相反，在全

服务链体系中，是以顾客的需求为供给出发点的，按需生产，精准供给。纵观产供过程，最为理想高效的方式是供给和需求之间能够零时差同步，这依赖于对真实需求的预测计算、对生产过程的合理规划。融合了庞大数据基数、响应高度灵敏、供给传输精确化的模块化网络需要做到以下几点：一是根据行为导向进行预测分析，对规律化的运作方式做出预测，并先行调配；二是增强信息传导速度，将传递时间压缩到最短；三是保持同步运作，增强响应能力，压缩滞后产生的时间差；四是提高协作能力的效率，增强模块组合连接的灵敏度和活性。

第四节 块数据全治理链

（一）数据感知和治理创新

块数据为国家治理体系和治理能力现代化提供了一种革命性工具，通过对数据的全面感知、采集、处理、分析，为政府治理模式创新带来新机遇和新动能。

治理体制从碎片化向融合化转变。当前，政府治理体制存在的突出问题就是治理碎片化，使治理实践陷入高成本和低效率并存的困境。碎片化的具体表现为：部门主义和地方主义盛行，职能交叉和重叠，数据孤岛和“数据打架”现象并存。通过块数据全治理链，提升数据使用效能、效率和效益，可以有效破除壁垒，实现治理数据的开放共享，推动多元治理主体之间的合作与协同，进而促进治理体制从碎片化向融合化转变。

治理方法从样本推断向“用数据决策”转变。传统占主导地位的治理思维是：根据部分人的治理需求推断和预判其他多数人乃至整个社会当下或者未来的治理需求，根据部分区域的治理经验推断整个地区乃至国家的治理政策和措施。这种类似于样本抽样推断总体的治理思维越来越显现出其局限性，主要表现在：一是未能充分考虑到社会各群体之间治理需求的差异性；二是对不同地区之间发展水平不平衡的基本国情缺乏解释力。块数据的价值在于提高决策科学化与治理精细化。各治理主体必须对原有的治理方法进行变革，着力强化数据治理意识，更多地通过价值数据进行决策，实现向“用数据决策”转变。

治理模式从静态向动态转变。社会治理的目标之一就是维持稳定，表现在治理实践中就是一种静态的治理模式。全球化、信息化和网络化，急剧改变着经济社会的变迁速度。面对许多新问题新挑战，块数据可以帮助治理主体及时全面地掌握相关数据的变动情况与趋势，促使静态的治理模式向动态转变。

（二）块数据全治理链体系

全治理链协调多元利益相关方，从提升治理能力和发挥治理体系效能两个方面入手，通过完善共治、善治、自治、德治和法治这5种能力建设，构建治理结构体系、治理功能体系、治理制度体系、治理方法体系和治理运行体系这5种治理体系，以块数据为价值中枢，以数据流动规范和优化信息发布、决策流程和绩效评估等环节，形成用数据说话、用数据决策、用数据管理、用数据创新的治理新格局。

从全治理链的驱动机制看，治理体系和治理能力是一个有机整体，推进治理体系的现代化及增强治理能力，是同一政治过程中相辅相成的两个方面。有了良好的治理体系，才能提高治理能力；反之，只有提高治理能力，才能充分放大治理体系的效能。块数据全治理链可以分别或同时从发挥治理体系效能和提升治理能力两个方面入手，带动整个链条的升级。

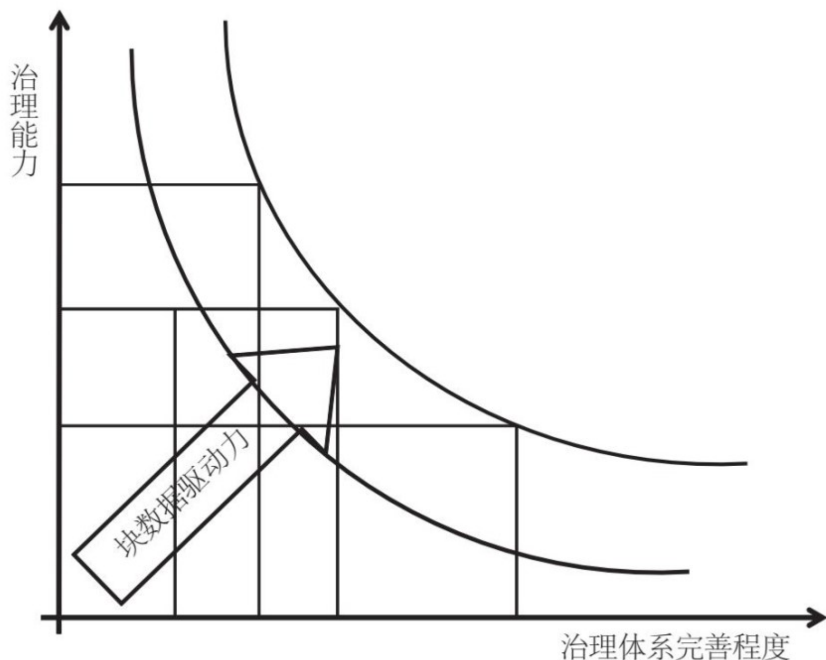


图6–15 块数据全治理链的驱动机制

通过块数据全治理链推进治理体系和治理能力现代化，一方面，要突出治理结构体系、治理功能体系、治理制度体系、治理方法体系和治理运行体系这5种体系构建；另一方面，要完善共治、善治、自治、德治和法治这5种能力建设。这5“治”，并不是孤立的、零碎的，而是有机的、统一的整体。共治是多元主体通过协商民主等手段发起集体行动以实现共同利益的过程；善治则是强调公共利益最大化的管理过程；自治是基层民主的重要实现形式；法治思维强调以权利义务为中心；德治则是确立有利于社会发展的社会价值观基础，以建立有利于国家强盛与人民幸福的社会隐秩序。“五治并举”是系统治理、综合治理、依法治理、源头治理的具体表现形式，是块数据全治理链构建多元参与、良性互动治理格局的核心环节。

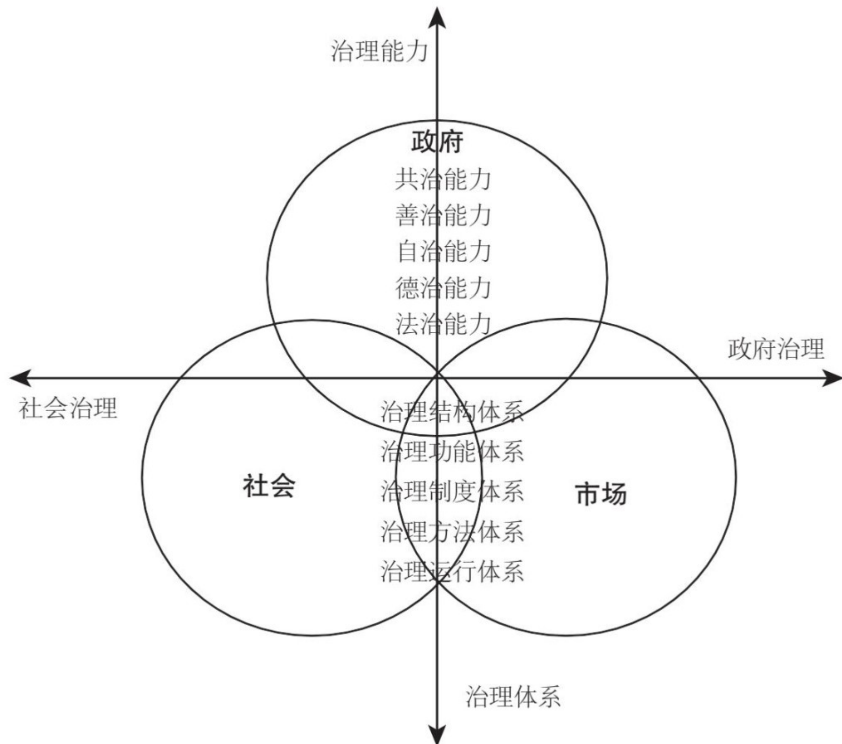


图6–16 块数据全治理链的结构体系

（三）块数据治理的精细化和精准化

通过全治理链，将推动政府治理由直觉经验决策模式向数据驱动决策模式转变，实现政府治理与治理能力的精细化与精准化。耶鲁大学埃斯蒂教授指出，“数据驱动决策”将使政府更高效、开放和负责，更多地在事实基础上做出判断，而不是主观判断或者受利益集团干扰进行决策。

人格数据化。全治理链模式中，块数据通过对每个个体人格的数据化，塑造了其主体性公民身份，这一身份认同流动于治理链不同层次和不同环节的数据集合体之内和之间。数据化的人格更容易与其周围的人、事、物建立连接，进而促进多元价值主体共同形成价值链网络结构。在这个价值网络中，没有单独的中心权威，决策的关键成为多元主体间的共同经验、学习过程和话语赋权。块数据全治理链中，更多的跨越传统边界的参与主体加入到公共决策过程中来。公共决策的目标更加强调覆盖群体的多元性、复合性和整合性，推动公共事务治理过程的民主化决策，实现多元共治的目标。

预测性决策。缺乏对现实信息的全面认知，是造成传统决策经常滞后于决策需求的主要原因之一。基于块数据的公共决策过程，可以有效实现决策的前瞻性、战略性、操作性，也将之带入“预测性决策”的新时代。块数据价值中枢所提供的分析和预测能力是公共决策的重要因素，通过价值数据流实现对要素资源的引导和配置，最终表现为政策执行结果的最优化。全治理链将从全局性角度，通过块数据实现围绕个人、法人、社会和国家的块上聚集，构建覆盖地理、时间、人口、环境、安全等领域的数据整合渠道；分析个体型数据、群体型数据之间的内在联系，通过块数据价值中枢发现数据集之间的关联、因果结构以及行为模式，构建预测模型；分析事件原因，标明潜在引爆点，进行行为预测并制订应对方案。

从预测性决策的具体分类看，可以分为回溯型数据决策、预测性数据决策和预置性数据决策[注]。回溯型数据决策可以理解为是“过去→未来”，强调对历史数据的分析处理，发现隐藏规律并做出预测；预测型数据决策是“过去＋现在→未来”，强调综合历史数据和现在的数据，通过构建仿真模型做出预测；预置型数据决策是“现在（激活条件）＋未来（达到条件）→未来”，强调通过对实时数据流的监测与分析，在预置激活条件被触发时完成自激活程序，提供相关预设决策方案。

量化政府绩效评估。政府绩效评估是块数据应用的重要领域。随着数据感知、数据采集、数据存储水平的进步，让原本不可识别的数据如身份数据、行为数据等变得可以被识别，许多原来不能量化的评估内容变得更容易量化，这大大拓宽了评估的范围和内容，也让构建更加精细的政府绩效评估体系成为可能。例如美国政府的“绩效仪表盘”项目，就是通过更大范围的数据评判构建的政府绩效评估机制。“绩效仪表盘”涵盖了对美国政府机构的绩效、联邦政府资金使用情况、政府资金使用情况、政府法律法规制定流程等多个方面，从不同角度对美国政府绩效进行量化评估与追踪。[注]

此外，块数据全治理链还可以对决策执行的过程和结果数据进行动态评估监测。治理链的网状结构简化了监督数据反馈的传输渠道，降低了反馈失真的程度。由于数据感知的范围覆盖广泛，可以对包括传统媒体、新媒体的媒体数据流，智能硬件的感知数据流进行综合汇总和研判，对存于其中的公共需求集合进行挖掘分析，找准痛点，变革决策反馈机制，进而不断优化评估考核体系。

第五节 块数据价值链的关键因素

（一）数据资源的开放共享

数据流是块数据价值链的核心资源。数据流之于块数据价值链就如同石油之于工业社会一般，是块数据价值链中的血液，只有充足顺畅的数据流循环再造输送“能量和养分”，才能够保证整个块数据价值链体系的正常运行。数据的流动带动了技术流、物质流、资金流、人才流、服务流在整个价值链中的合理分配和循环升级，将各行业、各领域的数据进行深度关联融合，进行系统性的整合、分析，发现新的突破点和增长点，产生价值放大效应。美国、欧盟等发达国家和地区已经认识到数据流能够产生的巨大价值，并认为数据的关联集聚程度将会成为衡量综合国力的标准，各国纷纷将数据开放上升到了国家战略高度。例如麦肯锡在2013年发布的《开放数据》研究报告所指出的，仅是教育、能源及医疗行业的数据开放，一年就能够为全球带来超过3万亿美元的经济潜力。

数据资源的开放共享是形成数据流的前提条件。数据共享和数据开放是两个有差异的概念。以政府数据共享和开放为例，国务院印发的《促进大数据发展行动纲要》强调要大力推动政府部门数据共享，稳步推动公共数据资源开放。这其中的政府数据的共享，指的是各级政府和不同政府部门之间的信息共享；而政府数据开放指的是将公共数据面向社会开放利用。

数据开放强调的是对原始数据的开放。但在目前的数据开放实践中，一些部门和人士还将信息公开等同于数据公开。其实，政府信息公开和数据开放的主要差别是开放对象不同。信息公开是开放到信息层，而数据开放是到达数据层。用一个比喻来讲，如果数据是面粉，政府开放了面粉，公众对面粉的利用可以是做馒头、面条、蛋糕……从而满足不同层次的需求，这才是激励底层创新的应有之义。但如果开放的是加工过的馒头，社会就无法再用这些馒头做成面条和蛋糕，那喜欢吃面条和蛋糕的人的需求就无法满足。在从面粉到馒头的加工过程中，数据中的一些原始价值或者被固化或者被剔除，因而失去了很多潜在价值放大和开发的机遇。数据是更底层的资源，只有充分的数据开放才能对社会产生更大的价值。只有对原始数据的充分开放才能形成包含更大价值的数据流。

数据共享重点是要破除数据孤岛和信息盲区，加快数据的交换与流动。目前对数据共享和数据共存的理解还存在误区，例如各地方政府认为建立统一的数据中心，将数据集中存储在同一个地方就能够打破数据孤岛，实现数据共享，而事实上这只是数据的共存，是一个相对静态的过程。共享强调的是不同层次、不同主体之间，对数据资源的分享和运用，而这种共享的主动性就是数据能够自由流通的内在驱动力，促使数据流在块数据价值链中的快速循环，实现数据价值的创造与提升。同时，高效的数据价值放大过程也能够反向促进数据资源的开放共享，在价值放大和数据资源开放之间形成良性的反馈回路。

（二）统筹规划的信息基础设施

信息基础设施是数据流的重要载体。数据流的形成涉及数据的采集、传输、存储、分析和应用等各环节，每个环节的实现都需要有效载体，而涵盖所有节点的载体集合就是信息基础设施。信息基础设施可以分为硬件设施和软件设施两大类，即信息技术所涉及的物理设施和基础网络系统。信息基础设施中最为关键的部分是数据采集平台、数据存储平台和数据应用平台，对应解决了块数据价值链中最为基础和核心的三个问题，即“数据从哪里来、数据放在哪里、数据谁来使用”。

数据采集平台。块数据的本质是连接，其本质就是要通过连接来构建个人数据、法人数据、社会数据和国家数据。如何才能实现连接呢？将相关主体的关联数据采集起来，并实现管理，便是实现连接的重要途径之一。随着数据采集设施的增加和能力的增强，能够获得的数据也不断增加。从某种意义上说，现在最缺乏的，不是数据，而是一个统一的数据采集平台。一般而言，大数据采集平台可以分为两个层次，即数据采集层和预处理层。采集层负责采集各种来源的数据；预处理层负责对采集回来的数据进行一些规范化的处理。建立具有强大能力的数据采集平台是信息基础设施建设的核心内容，它能够快速、高效、全面地提取相关信息、剔除冗余，得到正确、标准的数据，并具有一套完整规范管理体系。

数据存储平台。数据存储平台作为数据采集的上层系统，作为数据分析应用的支撑系统，也是信息基础设施的重要组成部分。数据体量的增加以及数据种类的丰富，这些新的情况对数据的存储提出了严峻挑战。面对挑战，利用云计算相关设施，可以提升访问和使用速度。基于

云计算技术构建的数据存储平台可以满足对结构化数据、非结构化数据和半结构化数据的统一存储和查询的需求。查询的高效性和存储的安全性是数据存储系统的两项关键指标。

数据应用平台。块数据价值链强调通过对数据流的整合、分析实现数据价值的放大。块数据真正的魔力在于应用，要通过政策和市场手段，鼓励和吸引更多市场主体加入到数据应用平台，推进创意与资本结合，建立数据开源机制，以开放理念助推数据应用。通过商业化应用模式的创新，让数据有价值、可交易，并产生红利。

信息技术设施建设要注重统筹规划。建设过程中要强调市场和政府的共同参与，相互协同。一方面，做好科学规划，加强顶层设计，体现政策有效性与调控效果，完善整合支持渠道，综合运用多种手段发挥政府的引导作用；另一方面，发挥市场在资源配置中的决定性作用。同时，要充分利用现有的数据资源和平台设施，对现有的数据中心和服务器进行改造和利用，构建布局合理、低成本、高效率的数据存储体系，避免重复建设和资源浪费。

（三）标准、政策和制度体系

数据标准体系的建设是实现块数据价值链的关键。由于数据的来源不同和数据主体的多样性，在块数据应用中出现严重的不同步现象，影响了块数据价值链的形成。解决这一问题的有效措施就是建立和完善统一的标准体系。块数据价值链运行的各环节都应该有相应的标准以保证每个环节的有序进行，如数据开放标准、数据采集标准、数据存储标准、数据交易标准等。在数据标准体系的建设过程中，政府应该统筹协调，开展标准验证和应用试点示范，建立标准符合性评估体系，并积极参与相关国际标准的制定工作。

完善相关政策体系是块数据价值链的重要保障。必须出台相应的政策，不断完善组织实施机制、健全市场发展机制、加大财政金融支持和加强人才培养，为块数据价值链的实现提供保障。

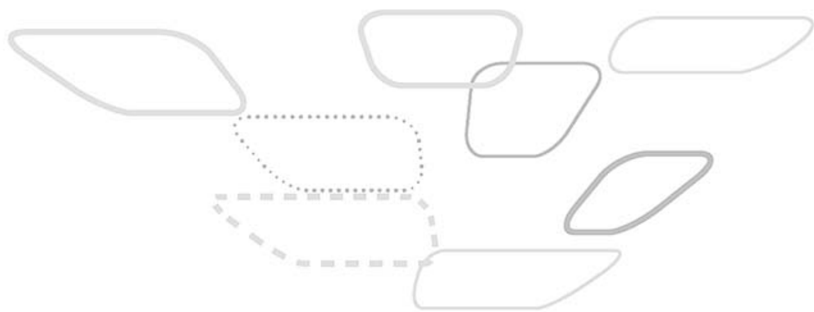
从法规制度层面来保障块数据价值链的正常运行也非常重要。政府要积极制定和完善数据开放和数据安全方面的法律法规，完善监管体制机制，在保证数据安全风险可控的前提下，推动政府、企业（尤其是数据垄断企业）最大限度向社会开放数据资源，实现块数据价值链中数据

采集、存储、利用和传输的安全有序开放，为块数据价值链的价值提升提供保障。

---

1. 汪辉勇：论价值的本质特征及其定义，《湘潭大学社会科学学报》，1999年第6期。
2. 车品觉：《决战大数据》，浙江：浙江人民出版社，2014年。
3. 迈克尔·波特：《竞争优势》，陈丽芳，译。北京：中信出版社，2014年。
4. 李长岭等：信息与知识价值链，《图书与情报》，2004年第3期。
5. 李健：知识价值链研究现状分析，《情报杂志》，2012年第2期。
6. 袁青燕：《价值网的竞争优势形成机理研究》，江西财经大学论文，2013年。
7. 中央经济工作会议要求准确把握经济发展9方面新常态，人民网，2014年12月11日。
8. 曲泽静等：新常态下价值链升级的创新驱动系统研究，《技术经济与管理研究》，2016年第1期。
9. 胡税根等：基于大数据的智慧公共决策特征研究，《浙江大学学报》，2015年5月。
10. 胡税根等：基于大数据的智慧公共决策特征研究，《浙江大学学报》，2015年5月。
11. 唐斯斯等：以“数据治理”推动政府治理创新，《中国发展观察》，2014年5月。

第七章
块数据组织



块数据既是一种经济模式，也是一种技术革新，更是一种新的世界观、价值观和方法论，引领和催生新的组织模式。在块数据组织中，传统意义上的资源稀缺已经被交互共享所打破，物质资源不再是阻碍创新的壁垒，人们不必再像过去那样，必须进入一个固定的组织才能获取所需要的生产、生活和创新资源。物物互联、业业互联、人人互联，构成一个自组织、自激活的数据实时流动、共享、聚合的数据化网络生态圈。块数据组织是一个资源共享、平台无限扩展的超稳定的组织结构源。人人组织、无边界组织、自组织、云组织、指数型组织等，都可以看作一种正在萌芽和生长着的块数据组织。块数据组织预示着组织发展的新方向。

块数据组织的出发点是数据人假设，利他主义的数据文化是构成块数据组织的理论基石。块数据组织中，数据力上升为组织的核心竞争力。数据力与数据关系影响着社会关系，这意味着数据力的改变将推动数据关系的改变，而数据力与数据关系的改变，又将引发整个社会发展模式前所未有的变革和重构。

块数据组织中，数据对组织的影响已经不再停留在技术层面，而是在战略层面为组织发展提供了全新的方法论。数据资源已经上升为一种基础性战略资源。资源的配置过程转变成为一种被数据化的过程，组织变得更加开放和更便于共享，组织外部协同成本比内部协同成本更低。组织领导力向平台领导力转型，是组织决策的根本性改变，组织的最高

决策者不再关心权力和地位，而更关注与外部的联系，走向一种“轻管理”模式。所有这一切，都预示着块数据组织“扁平化、平台化、关联度和聚合力”三位一体所带来的强大组织势能，实现组织的自激活和自适应环境变化，是组织存续和发展的根本动力所在。

第一节 数据人假设

任何一个组织，它的核心都是人，人是组织共有的资源和存在的基础。无论中国的传统文化还是西方的管理思想，都是建立在对人的认知和假设基础上的，如中国儒家的性善论、法家的性恶论、墨家的可塑论、道家的道性论，西方近代管理学思想提出的经济人假设、社会人假设、自我实现人假设、复杂人假设等。组织管理作为人类最基本的实践活动之一，在社会生产生活中起着越来越重要的作用。任何管理活动都离不开人，因此在管理中对人的研究是必不可少的。而研究人，必须通过假设，才能推导出组织的行为范式，建立起组织的基础模型，使组织理论走向科学。人性假设理论作为管理学中对人的认识，已经成为组织管理理论研究的出发点。[注]那么，块数据组织中，我们该如何定义“人”呢？

（一）从经济人、社会人到数据人

经济人假设。亚当·斯密最早在《国富论》中提出了“经济人”概念，是西方资本主义形成和初步发展时期经济社会实践在思维和理论层面的提炼。正是基于这一假设，人们发现，在一定的社会秩序下，个人追求自身利益最大化的同时存在着“看不见的手”，它能无意识且有效地增进社会公共利益。这一假设及相关理论既改变了当时放任自流的管理状态，又加强了社会上对消除浪费和提高效率的关心，更促进了科学管理理论的建立。

社会人假设。20世纪30年代，美国哈佛大学教授梅奥基于他的霍桑工厂试验提出了社会人假设，依据社会人假设形成的管理理论，使管理学从以物为中心走向以人为中心。[注]社会人假设强调人与人之间的关系和组织的归属感比经济报酬更能激励人的行为。社会人在生活工作中不仅有追求经济收入的动机和需求，还有获得友谊、安全、尊重、归属等情感的需求。但是，社会人假设是一种依赖性的人性假设，虽然它

意识到非正式组织的作用，却把这种非正式组织视为组织的对立，因而没有把这种基于情感联结的组织视为一种价值。

数据人假设。随着互联网和大数据时代的到来，数据爆炸正在驱动新的组织与共享模式，微博、微信、团购网站、社交媒体等成为新的瞩目焦点。这是一个人人时代，是一个数据人的时代，更多的组织将是一种基于话语的、临时的、短期的、当下的组合。数据改变了人类社会的沟通和认知方式，在未来块数据组织中的所有人和物都将作为一种数据而存在，作为一种数据而联系，作为一种数据而共同创造价值。与其说电子邮件、聊天室、博客、开放源代码等是聚集人气的地方，不如说它是一个协同合作的组织共享空间。价值与智慧就隐藏在数据之中，块数据组织让“数值想象力”的开发成为一种价值导向，我们称之为“块数据主义”。

数据人假设强调人的行为关系和存在方式的利他化。人作为客体被接入了互联网，成为一个不断采集数据并向云端传输数据的节点，开启了人的数据化。人的存在形式、生活习惯、思维意识等都可以由数据表达，同时可以通过共享、交易等方式获取任何其他人和物的数据表达。组织获取数据和产生数据的能力不仅大幅增强，而且实现了跨界关联，这种能力包括分享的能力、与他人特别是不特定关联人合作的能力，以及采取集体行动的能力等。

数据人不仅仅是指人，所有的物件和部件也都作为一种数据化的个体而存在并交互影响。条数据是围绕着物产生的，块数据则是以人或组织为起源的，呈现出主体性特征。既然人可以是数据化的，同样，物在本质上也是数据。一辆汽车，因为各种打车软件而化身为一个流动的数据，奔跑在城市的洪流中，其所汇聚的海量数据足以让数据工程师发动它所有的潜力。一部手机，就是一部新新人类的组织关系史。它就像蜜蜂的蜂房，既是一种社会性装置，也是一种有关蜜蜂的信息工具，它为蜜蜂提供了沟通和合作的平台，使蜂群的生存成为可能。人类网络也是如此，蜜蜂创造蜂房，我们创造手机，手机也是一种数据人的存在，是一个组织的元素，一个可以共享的、公共的大数据场域。

数据人假设的提出，超越了传统的善恶边界，打破了限制组织有效性的传统桎梏。大规模的业余化超越了传统组织利益关系的局限，大规模的自组织和自流程化降低了传统组织中的交易成本，他们的工作无须

遵循传统的管理规则，从而使利己与利他的有机统一成为可能。

（二）利他主义的数据文化

组织是一个有机生命体，它能够通过人格产生和传递组织的价值观和人格意识。块数据组织的人格就是一种利他主义文化的传播。19世纪法国哲学家、伦理学家奥古斯特·孔德首次提出“利他主义”一词，此后社会生物学家爱德华·威尔逊将利他主义分为两种类型：一种是纯粹的利他主义，是不求回报地增进他人的利益；还有一种是互惠的利他主义（即有条件利他主义）注。利他主义具有使他人得益的行为倾向和价值主张，是一种自觉的外化实践过程，能够增强个体的共享意愿，从而促进个体的共享行为注。

阿里巴巴董事局主席马云在一次会上说，现在这个世界要用好DT（数据技术），其核心就是利他主义，“相信别人要比你重要，相信别人比你聪明，相信别人比你能干，相信只有别人成功，你才能成功”。阿里巴巴不是一家电子商务公司，而是一家帮助别人经营电子商务的公司，要想成功，就要先利他，利他之后才能利己。注

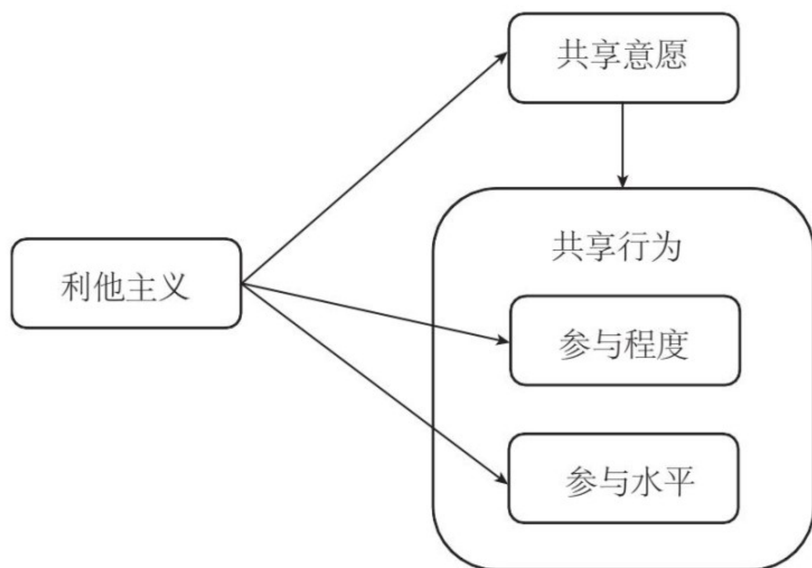


图7–1 利他主义与共享行为的关系

“利他主义的数据文化”最典型的例子是维基百科。在吉米·威尔士和拉里·桑格启动这个系统之时，并没有明确的指导思想。但理念、互联网和维基软件融合在一起，触发了全球知识共享的狂欢。为什么人们愿意

在一分钱都挣不着的情况下为维基做贡献？用经济学的话来说，就是因为它用开放式的社会系统戏剧性地降低了一般行政费用，也减少了对于参与的各种阻碍。正因为开放性实现了零边际成本，没有利润和所有者，而以低成本向人们提供了大量信息，实现了陌生人间的互惠互利。这一免费积累而成的庞大的知识库，不仅对未来的商业交易成本具有重要的含义，还对整个人类社会的思维模式和行为范式都产生了根本性的颠覆。相比《大英百科全书》200多年的权威历史，维基百科诞生不过十几年，但是后者无论是在开放性、信息量，还是迭代速度方面，都非前者可比。维基百科的出现证明了一个不争的事实：只要有机会，人们就希望分享他们所知道的，并且愿意免费这样做。传统权威走向了终结，我们迎来了一个信息自由分享、自由传递、自由呼吸，从“笼子里”释放出来的时代。

利他主义的数据文化已经成为大数据时代影响整个社会的一种主流文化，为社会的发展提供了源源不绝的动力和能量。亚当·斯密在《道德情操论》中指出：“如果一个社会的经济发展成果不能真正分流到大众手中，那么它在道义上将是不得人心的，而且是有风险的，因为它注定要威胁社会稳定。”[注]如果文化不止损，那么经济的止损效果就是有限的。利他主义的数据文化终于跳出了带血的资本积累的原始冲动，让经济发展的成果通过数据化的组织和共享得以分流到大众手中，这不仅表现在它所倡导的利用群体智慧进行的大众创新和万众创业之中，而且还表现为那些在透明的数据环境下诸多“独角兽公司”[注]所享受到的快感和惬意。

（三）从团队到平台

团队是组织理论的一个核心概念。传统组织理论认为，在变化的环境中，团队比传统的部门结构或其他形式的永久构成更为灵活，反应也更迅速。由于团队能够被快速地组合、配置、重新聚焦和解散，当组织为了更加有效果和有效率地应对竞争而进行结构重组时，团队就成为一种能够充分利用成员才能的最佳方式。团队之所以在组织理论中盛行的另一种解释是，团队是管理层增强组织中的民主气氛、提高积极性、实现组织目标的一种有效手段。组织围绕着团队来重新建构行动流程，通过团队的积极协调作用提高组织绩效。因此，在以往的组织中，团队建设是加强组织工作的主要方式。

但是，在块数据组织中，团队的地位将被平台所取代，平台属性相比团队变得更为重要。这是因为，传统的组织是串联的，自上而下的，而块数据组织是并联的，互联互通的生态圈，围绕共同关心的价值成果形成每个人自由发表意见的平台。平台实现了即时组合，同时促进平等、紧密的沟通和组织形态的建立。正如海尔所提倡的“外去中间商，内去隔热墙”，通过创建开放的平台，让所有的一流资源无障碍进入，给组织带来无限生机与活力。

平台让组织的价值最大化，每一个组织成员在平台上贡献着自己的智慧、资源和能力，并快速地组合着成长要素。组织成员不再固化在各自的岗位范围内，而是保持互动合作和高度自治。另外，在块数据组织中，平台成为资源开放共享的场所，这个场所可以是物理空间或虚拟空间，也可以是环境或者其他某种工具，资源要素也通过在平台上的关联聚合实现价值最大化。

从块数据组织中平台的特点看，平台是联结和交互的场所。在平台上，联结是基础，交互是方式。联结意味着拉近了距离，紧密了关系，意味着有共同兴趣、关注点和利益，在此基础上构成联盟和利益共同体。交互产生数据、知识和价值创造，信息传递不再是单向单维的被动交互，而是双向多维的主动创造。传统组织的串联型价值链变为同步并联的价值链，各个节点都共同响应并且在全流程中共享信息，彻底打破了命令式的结构，变成了自驱动的平台。因为联结，组织淡化了内外的区别，任何组织和个人都可以在平台上集聚。因为开放，大量的跨界行为涌现。例如，粉丝人力资本管理，通过粉丝交互平台可以挖掘大量的潜在需求；粉丝是创意的源泉，粉丝的深度参与能给迭代创新提供智慧的火花；最后，粉丝还是品牌的代言人，尤其是在自媒体时代，通过粉丝的口碑营销可以迅速提高产品的知名度，吸引更多的用户参与进来。

平台是自演进的有机更新体系。平台能够实现组织单元间共赢共享的市场交易关系，为组织提供优于社会的资源支持，如果平台不能满足需求，组织便可以不使用内部资源平台提供的服务，转而选择生态平台上更优质的资源，这就激励平台持续地追求卓越，提升竞争力，促使平台不断迭代优化。

更重要的是，平台本身具有风向标、领导者的作用，引领着行业和组织的发展。平台实现了“去中心化”的领导，是对中心权力的分布，不是不要中心，而是将一个大集权中心分为很多个小中心，那些占据领导

地位的平台将在更大程度上引领企业、行业和组织的发展。例如，目前出现的大多数“独角兽公司”就是通过做平台而实现高估值，它们都具有引领组织发展的领导地位，是平台领导力的杰出代表。

第二节 块数据组织的超稳定结构

组织结构是组织运行系统的框架，是组织内部进行价值创造的秩序，而有序性是秩序的基本要求。块数据组织是一种超稳定结构，表现为其具有自我调节的修复机制，能够进行内部协调和平衡，实现从混沌到有序，并具有混沌与有序的双重特征。在面对组织脆性和强控制的时候，块数据组织能够吸收和应用无组织力量平衡组织的结构，以无组织的组织力量进行修复，同时激活个体和组织的活力，是共享型组织的新范式。

（一）组织结构的演化

人类社会的组织结构经历了长期演化，日渐多元，包括科层制组织、扁平化组织和网络状组织，从简单到复杂、从垂直到水平、从封闭到开放、从有形到无形，变化无穷。当组织规模较小、价值创造活动较为简单时，权力集中在不同层级的管理者手中，科层制组织是一种有效率的组织结构。随着规模的扩大，为应对价值需求的变化以及解决金字塔式组织结构僵化的问题，组织逐渐转变为横向的运行秩序，开始重视组织价值的创造能力。组织朝着灵活性转变，从而出现了扁平化组织。

当组织实现了扁平化之后，组织会变得更加开放。纵横交错的价值创造链条构成了网络状的组织结构，组织结构更加复杂，组织形态也更加灵活，更能够应对环境的变化和不确定性，更具有开放性。开放性是一种生命力的象征，是组织存续的保证。日本女作家盐野七生在其名著《罗马人的故事》注中曾经做过这样的比较，公元前3世纪左右，古罗马在修路，而秦帝国则在修长城，结果到最后，罗马修了15万公里的路，秦帝国修了5000公里的长城，这反映出两种截然不同的思路：修路是沟通、开放；修长城是封闭、保守，其实最后也守不住。

网络状组织成为组织形态的主流。从演变的规律可以发现，“当一种组织结构向另一种组织结构演变时，往往具有两种组织结构的特点，属

于一种混合型的组织形式”。注 块数据组织是顺应数据经济时代而产生的复杂组织形式，是网络状组织形态的更高层级组织形式。同时，块数据组织边界超越一般组织的边界，不仅具有可渗透性和模糊性，还具有自相似、自学习与动态演进特征。控制论创始人诺伯特·维纳说过一句名言：“社会通信是使社会这个建筑物得以黏合在一起的混凝土。”注 这里的“通信”是广义的，当然也包括数据，因此，由于更广泛的数据共享，组织向环境学习，提高自身有机性与适应性，推动组织更好地进行环境匹配、协调组织与环境的关系，使其能够在复杂、不确定的环境中更好地生存和发展。

块数据组织正在崛起。块数据组织不仅是一个概念，而且是一种思维方式和一套方法论。例如，谷歌董事局执行主席施密特认为未来组织的关键职能是让一群“聪明的创意人”聚在一起，快速地感知客户需求，愉快地、充满创造力地开发产品、提供服务。而在此过程中，组织不会对这些“聪明的创意人”进行管理，只是负责营造氛围。组织不再是通过建立组织壁垒的方式获得成功，而是更需要形成开放与合作的组织结构，让外界容易纳入，只有迭代的愿景和迭代的战略性思考，才能让组织更智慧、更具弹性，步履更轻盈。

（二）脆性、强控制和无组织力量的再认识

脆性是组织的基本属性。组织脆性是在一定的历史条件下伴随着组织的产生而产生的，是针对组织在内外部环境的不确定性与信息的不完全性的双重作用下，产生的脆性积聚所可能导致的组织崩溃的抗击打能力。例如，中国2 000多年的封建社会就是一种脆性较大的社会结构，但不能因此而认为它就是容易瓦解的。脆性只表示缺乏弹性，而结构是否会被瓦解，还取决于其保持平衡的能力。正是由于儒家的家国同构、宗法一体等观念的存在，中国封建社会结构的脆性和强控制得以形成，并形成了一种能够保持组织的超稳定社会结构的重要机制。

从组织社会学的角度看，任何组织都是脆性的，因为组织的条件非常苛刻。但组织的生命却是顽强的，组织生命系统具有控制和保持平衡的能力。控制论把这种自动保持平衡的系统，统称为内稳定器。生命系统为了保持体内脆弱的平衡，发展出了一套对平衡进行强控制的系统，例如人体内发展起一套高级神经液体调节系统对平衡进行强控制。注 组织和生命虽然有所不同，但它也是一个活的生命体，总是竭尽全力保

护自己的生存。因此，每一个在脆性平衡之上建立的组织，都必然有一套保护系统。这个保护系统，既包含有组织的力量，也包含无组织的力量，这两种力量在相互作用中实现组织的超稳定结构。

块数据组织本身也是一种超稳定结构，但其脆性被韧性所取代，强控制也因组织中个体的高度关联和更多数据资源的共享而变为你中有我、我中有你的共生关系，强控制变为强聚合。无组织的组织力量、自组织的修复机制相互叠加，形成组织脆性的强大的激活和修复能力。由此，块数据组织就成为一种前所未有地实现了内部协调和自我平衡的组织，其稳定的结构或者说是稳定的状态通常难以被打破，是比稳定结构更加稳定的结构。

（三）自组织、自激活与组织的修复机制

德国理论物理学家哈肯认为，“从进化形式来看，组织可以分为两类：他组织和自组织。如果一个系统靠外部指令而形成组织，就是他组织；如果不存在外部指令，系统按照相互默契的某种规则，各尽其能而又协调地、自动地形成有序结构，就是自组织”。[注]因此，自组织是一种进化的概念，是组织在遗传、变异和优胜劣汰机制作用下不断地进行自我完善和自我修复，是一种能够实现活力和秩序双收的组织形式，它将不断提高组织适应环境变化的能力。

块数据组织理论在强调自组织的同时，还强调组织的自激活。这是因为，在没有外界刺激的条件下，组织通常总是由“动态”趋向于“稳定”，而要从稳态重新转向动态，除了施加更大的甚至是破坏的外力之外，往往还要让自身始终保持活力和激情，即自激活。诺贝尔奖获得者普利高津在其耗散结构理论中认为，在开放的系统中，通过交换，系统把熵输出给环境，或从环境中引入负熵，从而保持系统内部的平衡。块数据组织本身具有较强的活力，它具有自我新陈代谢、自我修复、自我调节、自我控制的属性，同时还具有保持自身有序性和发展变化的活力。因此，块数据组织的平台可以通过自组织与自激活不断引进负熵，从而带来组织结构的转型与优化。用公式表示就是：

$$Sa+So>N$$

其中，“Sa”即自组织（Self-activating），“So”即自激活（Selforganizing），“N”即负熵（Negentropy）。

这个公式表明，封闭是一个组织走向衰亡的根源。能够实现组织长期存在的唯一途径就是减少熵增，实现熵减。办法就是实现系统的开放，加快组织的平台化，这也是块数据组织更强调平台而非团队的原因所在。块数据组织更强调激活，从而不断地与外界交互物质和能量，当这种能量交换达到一个临界值时，就实现了一种平衡的“相变”（质变），组织就将由原来的看似混沌的无序状态转变为一种功能上的有序状态。因此，组织开放的复杂性决定了它的非平衡混沌的必然性，即“混沌序”——宏观无序，微观有序。平台的开放性可以使其不断引进负熵，让组织从低级到高级，从混沌到有序。用华为公司创始人任正非的话说，就是要把那种因利益所在而对企业的“爱”耗散掉，而用奋斗者、流程优化来巩固组织的力量。

混沌理论对块数据组织的自组织、自激活意义非常重大。既然组织是一个复杂系统，那么面对未来的不确定性，即所谓的VUCA [动荡（Volatility）、无常（Uncertainty）、复杂（Complex）、模糊（Ambiguous）] 环境[注]，组织的运行就必须“足迹”（LFP）化。2012年11月美国总统大选前夕出版的《面对与隐藏》（*Confront and Conceal*）一书认为，奥巴马对无人机、网络武器和特种部队的运用发展就是一种“轻足迹”战略，旨在保护美国利益的同时避免“消耗美国人民的生命和财富的旷日持久的地面战争”。在块数据组织所面对的VUCA世界里，也体现为这样一种轻足迹的管理。比如，最新的机器人、自动化和数字化技术等，虽然它们本身并不是目的，但却是让组织变得更加轻巧、灵便的必经之路。而让组织足迹变轻中至关重要的，是以高度信任为纽带的合作与协同，以此产生自组织和自激活的动力。同时，这种自组织和自激活还将产生一种高度信任的模块化结构，组织及其模块能够自我调整，相信内部人也相信外部人，是跨界和开放的合伙人。进一步讲，块数据组织只用最合适的人，不用最聪明的人。最聪明的人永远在组织外部。

人人都是自己的CEO（首席执行官）。自组织、自激活破解了组织活力衰竭的困境，激发了组织自身的潜能，促成了组织中个体的自我完善以及自我价值实现，形成一个价值生态圈，同时推动“人人创客”战略的实现。这种个体激活同时催生了组织形态的变革，让组织变为共享平台、驱动平台和创新平台，推动了组织的去权力化、去中心化，实现管理的轻足迹化，让组织彻底跳出了兴盛与危机的周期律！

（四）共享型组织的新范式

让共享成为主义，这是互联网、大数据给人类带来的最大红利。英文中小写的“共享”（share），是个美好而古老的词，当它成为大写的主义——“共享主义”（Sharism）时，勾画出的是一个正在向我们走来的新时代——即“共享社会”（sharing/sharable society）新时代，它所开启的并非只是一种全新的商业模式，同时也是一场共享社会的变革，激荡着共创共治、无边界的共享组织新范式。

人与组织的关系从交换价值转变为共享价值。美国趋势经济学家杰里米·里夫金说过，未来社会可能不再是简单地交换价值，而是实现价值共享。过去所有的东西如果不交换就没有价值，但是未来不是交换而是共享。在块数据组织中，组织与人的关系发生变化，人与组织的关系不再是一种被动的、人去适应组织的关系，它完全颠覆了过去“资本雇佣劳动”的概念，是人力资本与组织资本、货币资本之间共赢、共享的合作伙伴关系，组织成为人开展价值创造活动的舞台，人也不再像过去那样被束缚，组织与人彼此依赖，或者说组织更依赖于人。例如，目前的华为公司已经实现了人与组织的这种共生关系，任正非对华为人保证公平原则、利益共享，甚至对华为的下游供应商们，都会在危难时期承诺“绝不让利益共同体吃亏”。

以价值创造为本，构建共融的价值生态圈。人本身的成就感是价值创造最重要的因素，只有以价值创造为基准来确定人在组织中的地位和作用，才能既尊重了人性，又避免了以人为本的陷阱。块数据组织的共享机理要求：一方面，解除原来对人的束缚和固化，释放人的价值创造主动性，激活人对价值追求的自驱力；另一方面，由于人人都是自主价值体，要以价值创造来分配资源。块数据组织在扁平化、自组织化、网络化之后，不再是一个小我的组织，而是一个大我的系统，这个系统体现为一种共享型组织的联盟，即联盟共存、联盟共创、联盟共治、联盟共担、联盟共享。整个系统组织与外部环境之间是完全开放的，是一个资源能够自由流动的系统，这种无边界的组织范式将让社会财富、资源得到极大的共享。而组织的价值已经不在于组织本身，而是在于其中每个个体的创造力如何被激发与再创造。谁能创造平台，让价值链成员众享价值，谁就会成为未来组织的领先者。

以共同治理形成利益共同体。意大利政治哲学家尼可罗·马基亚维利说过，如果不能使参与这件事情的所有人都能获利，那么这件事情就不

会成功，即使成功也不能持久。互联网、大数据时代之前，情况大抵如此，个人与企业、企业与企业之间，很大程度上都是零和博弈，此消彼长，不是你死就是我亡。但在块数据组织的新范式下，强调的是数据人的利他主义，是对经济人假设的超越。传统的价值分配制度下，价值创造与价值分配之间往往是割裂的，而基于利益共同体的利益最大化机制，价值创造与价值分享紧密地联系在一起，形成一个互补型生态系统。价值创造的目的是价值分享，以共享利益的方式驱动各方参与价值创造才能实现价值创造的最大化。

第三节 数据力：块数据组织的核心竞争力

生产力和生产关系是人类社会中最为重要的一对关系。进入大数据时代，毫无疑问，也存在数据力和数据关系的问题，而且这是一个值得深入研究的重大理论问题。从普遍意义上说，数据力是推动数据时代发展的根本力量。由于这种力量的相互作用影响，整个社会生产关系被打上了数据关系的烙印。这种新的数据关系，它关乎的甚至是如何重新按下资本主义死亡的“快进键”这样的大问题[注]，这是多么重大的理论发现。生产力与生产关系的矛盾，是推动人类社会发展的根本动力。数据力和数据关系何尝不在推动着数据社会加快发展。决定数据力高低的主要因素是数据人，而数据处理能力的水平则是衡量数据力水平的重要标志。我们现在要做的，就是顺应数据时代发展的内在要求，推动块数据组织、块数据经济、块数据政府的发展，使共享社会及其意识形态和上层建筑更加完善。

（一）从学习力到数据力

数十年来，围绕组织理论研究领域，各种学说不断涌现，各种理论交互影响，推动组织理论研究不断向纵深推进。尤其是20世纪90年代，美国学者彼得·圣吉《第五项修炼：学习型组织的艺术与实务》[注]一书出版后，“学习型组织”理论很快风靡全球，受到各界的重视与推崇，并成为组织与管理理论中的显学，许多知名企业纷纷转型为学习型组织，大力推行学习。学习型组织强调的一个重点，就是学习力。在学习型组织中，有一个非常著名的定律，即$LGC=D$，即当学习的速度小于或等于变化的速度时就等于死亡。这个公式影响了整整一代人，也奠定了学习型组织的历史地位。但块数据组织的定律，即$K=S(V,P)$，更强调

组织在特定平台环境下的关联与聚合，不断消除因为封闭组织文化、组织规范的固化产生的熵，从执行到自驱动，形成组织生命周期创新的持久活力。

表7–1 块数据组织与学习型组织的区别

组织类别	组织结构	组织资源	组织形态	组织优势	组织效能
学习型组织	稳定结构	人才	团队/合作	学习力	平衡/内部性
块数据组织	超稳定结构	数据	平台/关联	数据力	再平衡/外部性

块数据组织是未来引领组织发展和变革的一种新形态。与学习型组织理论相比，块数据组织把原来的组织学习引向了一个范围更大、模式更加开放的深度学习领域，推动了组织的学习转化，并由这种深度学习和深度转化形成了组织的新的核心竞争力——数据力。块数据组织是一种更高级、更具影响力的组织模式，是学习型组织的升级版。

数据力是大数据时代人类利用数据技术认识和改造自然的能力，它既是一种认知能力，又是一种发展能力，归根结底就是一种数据生产力。数据力是块数据组织的核心竞争力。正如前文所指出的，数据之间存在相互作用，这种相互作用是因为数据质点之间的数据引力形成的数据引力波。数据引力波将大量的零散、割裂的数据有机地关联起来，极大地释放了数据力的潜在价值。由于数据引力波的推动，组织之间可以实现完全对接，对数据的追本溯源将形成全链条数据力合力，将海量数据转变为直接的生产力，从而实现数据力的极大解放。

数据力的影响在今天已然显现。谷歌人工智能机器人AlphaGo战胜围棋冠军的影响力还在持续，充分展示了人工智能在未来领域的一切可能。10年前电脑击败了国际象棋冠军，说明电脑计算能力足够，但围棋的变化是指数级别的，人工智能战胜人类似乎完全不可能通过提高计算能力和穷局法来完成。但最终结果颠覆了所有人的预期，谁都没有想到，仅仅两年时间，基于大数据的人工智能的深度学习算法就取得了如此巨大的突破。可以肯定，用不了几年，智能化深度学习领域就会有各种应用产生，激励人才和资本大量涌入，进一步推动人机学习、数据互搏、数据驱动，从而在整体上提升组织的数据力。展望未来，数据力将是人类最重要的生产力，生产力将因数据力而空前提高。人工智能不会取代人类，但是人工智能将是人类数据力的集中体现，它将把数据力推

向极致。

（二）5种数据处理能力

数据处理能力是数据力的最重要组成因素，数据处理能力的水平是判断数据力水平的重要指标。未来组织的核心竞争力取决于它的数据处理能力。在一份2002年2月向媒体发表的声明中，时任美国国防部长，也是后来被奥巴马所接受并发扬光大的轻足迹军事管理的首席设计师唐纳德·亨利·拉姆斯菲尔德提出了著名的“三段论”：有所知的已知，有所知的未知，有未知的未知。在大数据时代，数据力和数据关系的相互作用以及数据处理能力的提升将能够帮助我们把所知的未知变成所知的已知，并找到未知的未知，将之变成所知的未知，甚至是所知的已知。

目前来看，大数据时代我们最缺乏的是对数据有用性和有效性的认知和挖掘能力，而不是技术，需要解决的是对数据的认知和跨领域的应用和处理问题。“数据处理能力是指会整理、分析数据，能从大量数据中抽取对研究问题有用的信息，并做出判断。”[注]数据处理能力的特点是用数据说话，说真实和有用的话，用数据穷尽真理。数据处理能力的目的是从数据中提取有效的、有用的信息，为做出决策提供科学有效的依据。这种能力包括数据采集能力、数据存储能力、数据关联分析能力、数据激活能力和数据预测能力。

表7–2 5种数据处理能力

5种能力	主要内容
数据采集能力	组织作为数据的使用者，能够掌握基本的数据获取方法，从适当的渠道获取所需要的数据；组织作为数据的生产者，能够通过记录、观察、实验等方法采集实践活动中所产生的各种数据；善于利用各种数据工具在互联网和其他数据源里采集所需数据；能够混合使用购买、搜索、截取等方法采集数据。
数据存储能力	能够以适当的格式对不同特点的数据进行保存；能够通过连接大量大规模、高容量的服务器存储数据；懂得运用“集群式网络附加存储”满足多节点数据访问和存储的服务方式；通过购买和租赁云储存供应商的数据存储容量存储数据；能够使用适当的数据存储工具对存储的数据进行全生命周期的科学管理，能够在必要时及时提取使用。
数据关联分析能力	能够准确理解数据的含义及其产生背景；能够按需利用关联规则分析法、回归分析法、归类分析法、社交网络分析法、自然演化算法等方法对数据进行必要的关联分析，找寻数据之间的逻辑关系，并能够把分析结果以最适当、最直接的方式呈现给需求者；能用辩证思维看待数据，不盲目崇拜数据，又能坚持用数据反映客观现实的原则。
数据激活能力	能够发掘抽象化数据、暗数据的潜在价值；能够以恰当的方式发布自己获得的数据；能够根据不同的受众及时采取相应的数据交流方式；能够预见性地进行自主搜索数据，在分析前期关联的基础上，更为精准地预判并搜集全面的数据资源。
数据预测能力	能对数据保持较强的敏感度，能够使用数据进行预测；能够使用科学的理论和方法建立数据预测模型；能够运用数据准确预测未来一段时间的行业表现以及发展趋势；能够运用数据进行更为准确的监测、预测、预警。

在这5种能力的基础上，块数据组织的数据处理还强调要通过人工智能、新型智库和数据科学家的加入强化新的数据处理平台建设。要利用人工智能进行更大规模和更深层次的数据处理，在更高层面强化数据的自组织和暗数据的自激活，尤其是要进一步发现数据引力波的作用和影响，推动数据力在更大范围和程度上的聚合和裂变。

（三）再平衡：块数据组织战略制高点

全球经济失衡是当今世界的重大难题，努力实现再平衡自然成为当

今世界的共同任务。同样，互联网、大数据时代带来的传统组织和块数据组织之间的失衡也持续性存在，所以需要在失衡中找到平衡点以实现组织的再平衡。这种再平衡的关键，是更加重视利用外部资源巩固自身的战略地位和可持续发展。变化就是存在。谁也不会想到出租车司机竟然是第一个全面与互联网对接的职业，没有想到原来每个人可以把自己闲置的车组合起来，通过一个非常便捷的平台租给个人，让车在空闲时能够帮助别人，同时又能够获取商业价值。不管你是否愿意，这就是现实，这就是再平衡的内在逻辑。只有外部联系才能打破传统组织的边界，才能让组织更开放、更扁平、更简化、更有效，让多元文化更加融合。这不仅是个人数据、平台和技术的再平衡，而且将实现个人愿望、组织目标和社会责任的再平衡。

再平衡推动“数据友好”型组织新愿景。复杂世界充斥着大量数据，以至于极难找出可以赖以决策的明确信息。就像给药的剂量超过限度治病的药就会变成毒药一样，当信息量极少时，每条数据都是有价值的，而当我们被大量信息淹没，其中大部分都是垃圾信息时，它就会让人无法选择，会造成摩擦和不信任，损害彼此的关系。就像二氧化碳，如果太少了，会没有植物，如果太多了，会产生蔓延全球的气候变暖，我们也无法生存。因此，组织面临的挑战，是要在两层意义上做到“数据友好”：必须从源源不断的数据流中提取具有价值的信息；需要快速地提取“切中要害”的信息。块数据组织所强调的“轻足迹”，就是要求组织要从大数据中挖掘具有价值的信息，更有效、更快捷地实现搜索、存储、过滤、提炼，或者如前面所说的高度关联和聚合，支持各层次的管理者做出及时、明智的决策，取得最大化的效益。

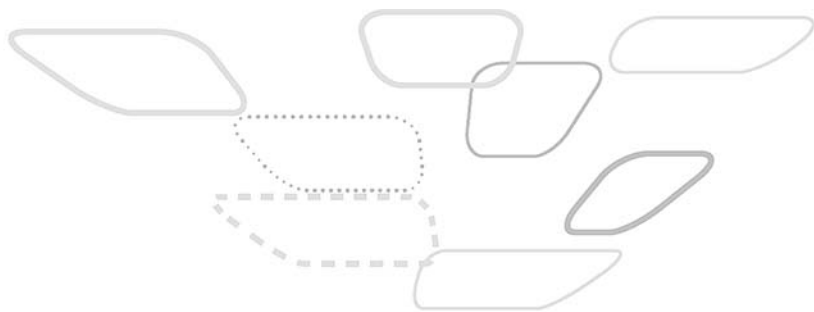
“离圆心越近，离失败越远。”在互联网时代，只有更具有动态平衡的眼光才有利于组织的茁壮成长。凯文·凯利总结的大自然生长的9条规律强调，要从无中生有到变自生变。也就是说，任何大型复杂系统都是协同变化的，只有在对称和均衡中才能形成一个安全的圆心，组织的平衡就要围绕这个圆心运转，在变化中实现动态平衡，从而远离风险。老子在《道德经》里讲“道生一，一生二，二生三，三生万物”，强调的也是一种从平衡到失衡再到新不均衡的过程。只有在动态的平衡中才能打破组织生命周期的魔咒，实现组织可持续和再平衡，从而在一个全新的平台上建立起未来组织治理的全新生命周期。

---

1. 张兵：哲学视野下管理学中的人性假设，《苏州大学硕士论文》，2006年4月。

2. 张建云、毛文龙：“经济人”与“社会人”的逻辑关系及当代意义，《经济研究导刊》，2009年2月。
3. Wilson, Edward. Sociology: The New Synthesis. Cambridge,Mass:Harvard University Press,1975.
4. 共享行为包括参与水平和参与程度，这两个参数能直接反映出个体的共享行为差异。从共享中互利，每个个体在社群和大众中都是一个获利者，同时也是一个贡献者。通过共享和互利，群体会变得更加和谐与可持续。
5. 许宁：阿里巴巴商业模式“三级跳”——马云C2B猜想，《经理人》，2012年9月。
6. 亚当·斯密：《道德情操论》，樊冰，译。山西：山西经济出版社，2010年。
7. 所谓“独角兽公司”，是指那些估值达到10亿美元以上的初创企业。
8. 《罗马人的故事》一书简体中文版已由中信出版社于2013年出版。——编者注
9. 杨少杰：《进化：组织形态管理》，北京：中国发展出版社，2014年。
10. 维纳：《人有人的用处》，陈步，译。北京：商务印书馆，1978年。
11. 金观涛、刘青峰：《兴盛与危机》，北京：法律出版社，2011年。
12. 姚文建：自组织理论下的国家开放大学办学体系建设探索，《中国远程教育》，2013年5月。
13. 张瑞敏：VUCA下的零距离和从零开始——《轻足迹管理》的三点启示，《商学院》，2014年12月。
14. 这是美国学者杰里米·里夫金于2014年由中信出版集团出版的新著《零边际成本社会》中的重要主题之一。可惜的是，国内的各个中译本都过滤掉了该书副标题中的“资本主义”（Capitalism）这个“大词”，但这恰恰是此书的核心立场，即“分享经济”（sharing/sharable economy，也译作“共享经济”）新时代，关乎的是“西方道路”这样的大问题。
15. 《第五项修炼》一书简体中文版已由中信出版社于2009年出版。——编者注
16. 李红梅：数据处理能力的内涵，《课程教材教学研究》，2014年。

第八章
块数据经济



块数据是推动新产业、新业态、新模式的强大引擎。块数据通过平台集聚各方需求，放大了创新、运营等各个方面的价值关联，实现了新科技革命和新产业变革的深度交叉和融合。这种融合充分利用块数据的渗透性和带动性，整合各经济体内部与外部资源，实现资源共享和资源互补，提升融合的深度和广度，从而形成催生新经济的新动能。

块数据经济是一种新的经济模式，具有资源数据化、消费协同、企业无边界、零边际成本、极致生产力等特点。块数据经济强调通过容错性创新试验，推动传统产业与大数据融合发展，深化大数据在全产业链的示范应用，引领着经济社会发展的高质量演进和系统性提升。痛客经济的提出是块数据经济理论的一个重大创新，对于重构创新产业链、重塑新经济形态具有重要意义，必将全面提升产业发展支撑能力和技术创新能力，实现个人价值、社会价值和经济价值的最大化。

块数据将解构和重构资源配置方式，从经济动能、经济结构上改变传统的生产力和生产关系，深刻地推动整个经济格局的变革和价值链的重构。这种重构将进一步改变社会关系，实现效率与公平的高度统一，从共享经济迈向共享社会，实现共享未来。

第一节 块数据与新经济

（一）资源数据化

资源是经济发展的基本条件。资源是稀缺的，所以人类的经济及一切活动永远面临资源的配置与选择问题。亚当·斯密在《国富论》中定义三大生产要素分别为劳动力、资本和土地。大数据时代，数据成为重要的生产要素。相比大数据带来的数据资源化这一变化，一个更具价值的趋势是资源数据化。资源数据化的本质是资源配置方式的变革，某种意义上说，块数据经济带来的是一种全新的、智慧的资源配置方式。

数据资源化		资源数据化
大数据的特征		块数据的特征
数据转化成资源		资源量化成数据
资源构成的变化		资源配置的变化
以竞争逐利为目的		以合作共赢为动力
资源整合出现新壁垒		资源获取实现零门槛
集中化的资源配置方式		平台化的资源配置方式

图8–1 数据资源化与资源数据化比较

动力机制：从竞争逐利到合作共赢。资源配置的目标是实现不同经济主体的利益，这也是进行资源配置的动力所在。在传统经济发展模式中，资产的所有人对其资产具有所有权，所有权具有排他性。为了满足经济主体的利益，排他性产权机制又再次激励持续投入，实现新一轮的资源配置。在块数据经济发展模式中，资源数据化使得资源共享成为可能，按照“但求所用，不求所有”的原则，所有权分离为支配权和使用权。这一点与古罗马时期的财产取向非常类似，通过以租代买，卖方也就是资源拥有者能获取一定利益，避免资源闲置带来的浪费，租方通过支付一定费用获得资源的使用权，这种暂时租赁使其能够通过资源投入满足需求，获取利益。传统的资源配置的动力是竞争与逐利，块数据经济的资源配置动力是共赢和多赢。

信息机制：从信息不对称到交互零成本。资源的合理配置有一个重要前提，那就是在及时全面地获取相关信息的基础上，形成合理方案，而信息的采集、传递、分析和利用需要有一定的渠道和机制。在传统经济模式中，经济主体通过信息不对称获得竞争优势，资源配置的信息传递是有限的或单一的。虽然在互联网和大数据时代，市场信息海量喷涌，交易双方可以通过信息技术和网络手段瞬间接触彼此，消除信息

不充分、不对称等问题，但同时，对海量复杂信息进行有效的汇聚、甄别、匹配和管控依然难以实现。通过块数据的发展，对数据进行采集、存储、关联、激活和预测，我们可以从海量数据中发掘隐含的关联价值，消除原有的信息壁垒，使获得精准信息成为可能。获得数据就等于获得资源，资源获取的门槛降低。

决策机制：从集中化到平台化。资源配置离不开决策，在传统的经济模式中，资源配置的决策权往往是相对集中的，通过集中的权力体系形成制约关系。大数据为人们搭建了一个又一个共享平台，平台充当双面角色，形成新的资源配置的决策体系，各个经济主体可以做出效用最大化的选择。比如，Uber的发展就得益于其为客户提供了一个闲置汽车的共享平台，将有闲置车辆的客户和有用车需求的客户联系到一起，实现双方的互利共赢。这种基于互联网平台的汽车共享商业模式不仅改变了人们的出行方式，更使得分散的个体能够基于平台实现共同决策，提高资源配置效率，并使得资源增值。这时，共享经济便出现了。共享经济提供了快速配置资源的平台，随着块数据的发展，共享经济将实现从平台到平台化的转变。平台不仅能够为经济主体提供资源，而且提供机制，提供流程，形成全新的创新体系。按照熊彼特的创新理论注，创新就是要“建立一种新的生产函数”，就是要把一种从来没有的关于生产要素和生产条件的“新组合”引进生产体系之中，这就是平台化的作用。

（二）协同消费模式

满足需求是经济发展必须面对的问题。从社会再生产的各个环节上看，与需求满足相对应的是消费，消费是最终的目的和动力。纵观人类历史，生产与消费的关系不断调整，并由此形成不同的经济模式。

生产者经济：生产必实现消费。在前工业社会，人类社会处于短缺经济时代，传统工场手工业生产占据主导地位。由于当时生产力水平发展低下，可利用资源相对短缺甚至匮乏，商品长期处于供不应求的状态，这就决定了市场长期处于生产者主导的卖方市场，消费者的需求偏好并未受到关注和重视，整个生产过程是：生产——消费——再生产——再消费……生产者经济具有“生产必实现消费”的特点，从本质上来讲这样的生产是无法满足消费需求的。在这种情形下，生产什么就消费什么，生产多少就消费多少，人们并不能根据实际需要和消费意愿进行

消费。可以说，生产处于完全的支配地位，决定了消费的对象和内容。

消费者经济：目标消费群体引领生产。到18世纪中叶，随着社会生产力的发展，产品极大丰富，卖方市场转变为买方市场，消费逐渐活跃，一个被称为“消费社会”的时代悄然来临，生产过程变成了：需求——生产——消费……由原先的生产决定消费的方式、质量和水平，变成了由消费所形成的新需要引导生产的调整和升级。“顾客就是上帝”口号的提出，就是生产和消费关系发生变化的集中体现。最具代表性的事件是，20世纪20年代市场营销理论在美国创立，企业开始通过市场调查、有关消费者心理和行为等的研究来调整生产。美国老牌汽车公司福特也由大规模的流水线生产，变为以满足用户个性需求为目标的小规模、小批量生产。这些都标志着“消费者主权”时代的到来和消费对生产的引领。

创造者经济：协同消费与产消合一。与生产者经济、消费者经济不同的是，创造者经济是一种以合作和互利为核心的经济模式，是随着协同消费的出现而出现的。协同消费的出现，得益于社会消费价值观念的变化、特定阶段经济形势的推动和信息技术的有力支撑。首先是观念变化。20世纪50年代开始，西方发达国家进入了一个高消费时代。但与此同时，人们也开始反思，炫耀消费、透支消费并不会让生活更幸福。人们开始强调借用、租赁和共享。[注]由此，作为美国梦根源的“财产所有权”受到冲击。其次是经济危机。2007~2010年席卷全球的经济危机，本质上是一种“左手闲置，右手短缺”的危机，急需某种商业机制进行调节。再次是技术发展。随着新一代信息技术的发展，办公共享、汽车共享、公寓共享等各种实体资源共享，通过互联网、物联网、大数据等虚拟互联网络得以实现，并产生一系列“波及效应”，掀起新一轮消费革命，也就是所谓的协同消费。

协同消费模式聚合需求、精准匹配供需的功能，更加契合消费结构的变迁和新价值诉求，已成为Web2.0[注]浪潮中的一种新文化，其不再是传统生产消费模式的附庸，而是产生了创造者经济模式。块数据经济就是一种创造者经济，其核心特征是产消合一。这是托夫勒在《财富的革命》一书中提出的概念，也就是生产者即消费者，消费者即生产者，生产者和消费者之间的传统界限已经模糊，生产、分享、销售、消费、再生产交织发生。消费者参与生产并主动参与创造，从经济链条的末端走到了前端，他们是具有生产性的潜在生产者。用户原创内容的爆发就

是消费者生产力释放的证明。这种产消合一的协同模式，通过再生产、再加工、再利用使资本创造和消费成为一个循环。在某种意义上说，块数据经济创造的就是消费资本化、产消联盟和价值共创。

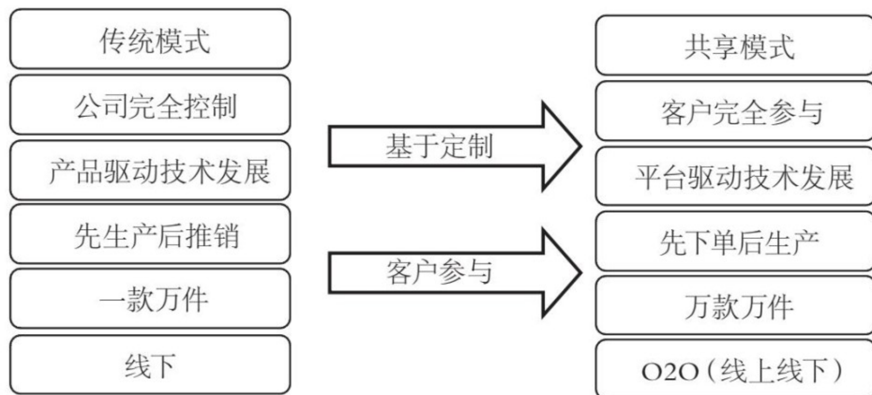


图8-2 传统模式与共享模式制造过程的区别

（三）零边际成本和极致生产力

2014年，杰里米·里夫金在《零边际成本社会》注一书中首次提出“零边际成本”和“极致生产力”的概念。他认为：“如果生产一件额外商品或服务的成本几乎为零，则代表着生产力处于最佳水平。”他更多的是从通信、能源和运输三大系统来论述这一观点，其关于成本与效率的研究视角，对于开展块数据经济研究具有参考价值。块数据经济将是一种成本更低、效率更高的发展模式，极致性体现在生产、服务、流通等各个环节。

从流程化到模块化，实现极致分工。在工业经济时代，信息采集、外包协作等成本很高，企业为了降低成本，更多的是从企业内部与行业分工入手，对生产、分销等环节实行垂直化和一体化的流程管理、链条管理，以此抵御外部市场的不确定性。随着信息技术、网络技术的飞速发展和广泛应用，传统的“大而全、小而全”的流程化管理越来越缺乏竞争力。由于资源在全球范围内的流动、配置、使用成为可能，设计、生产、分销、财务等功能可以实现模块化，突破有形界限，进行极致分工与协作，推动产业结构与产业体系重构，以此获得成本优势、创新优势、规模优势。

从重资产到轻资产，实现极致服务。目前很多企业都选择向轻资产转型，利用自雇型劳动者，实现更合理的企业劳动力结构，更加有弹性

地匹配市场高峰和低谷的供需，从而能够更加高效地响应服务需求。例如，Airbnb的平均价格与美国主要旅游市场酒店价格相比低21.2%，Airbnb在价格方面拥有绝对优势。维基百科是免费的，而《大英百科全书》的售价则高达1 400美元。这些C2C（个人对个人的电子商务）商业模式都具有轻资产经营能力，比起投资一家新酒店或者一个新的知识领域，其成本几乎可以忽略不计。更加重要的是，轻资产经营模式能够通过极致服务，拥有快速领跑市场的潜力和贴合服务需求的柔韧性，这是其成功的奥秘所在。

从点对点到多对多，实现极致价值。供应商和购买者在价值链上要真正建立起联系必须依赖于流通驱动。目前，网购已经成为人们的一种生活方式，不再是一种点对点的流通。零售模式的颠覆仅仅是开始，智慧物流将更进一步地改变流通模式。人们用高速公路来比喻信息网络建设，今天我们开始采用大数据技术改造物流业，传统的点对点和中心辐射型物流让步于多对多的分布式联合物流。过去由一个司机负责全部载货的模式将被取代，分布式交付依托物联网实现物流接力。这种物流共享模式，优化了每个客户的物流需求，使得企业获得成本效益，并促进整个价值链条的价值最大化。

（四）从效率到公平

从经济学的角度看，效率是指在一定成本基础上所能获得的收益。公平是一个实证性概念，体现的是人与人之间的一种平等关系和状态。效率与公平是经济政策追求的基本目标，然而现实的状况是这对矛盾体总是难以协调，“鱼和熊掌不可兼得”。甚至还有一种“公平效率替代说”，认为如果效率和公平能够量化的话，一方的增加将带来另一方的减少，这种关系可以用减函数来表示。进入大数据时代，共享经济似乎找到了破解公平与效率悖论的方法。共享经济是一种协同消费模式，在一定程度上实现了起点公平，有效抑制了经济差距，改善了社会公平状况。块数据经济能够更加接近帕累托最优[注]的理想状态，促进效率与公平之间的平衡。

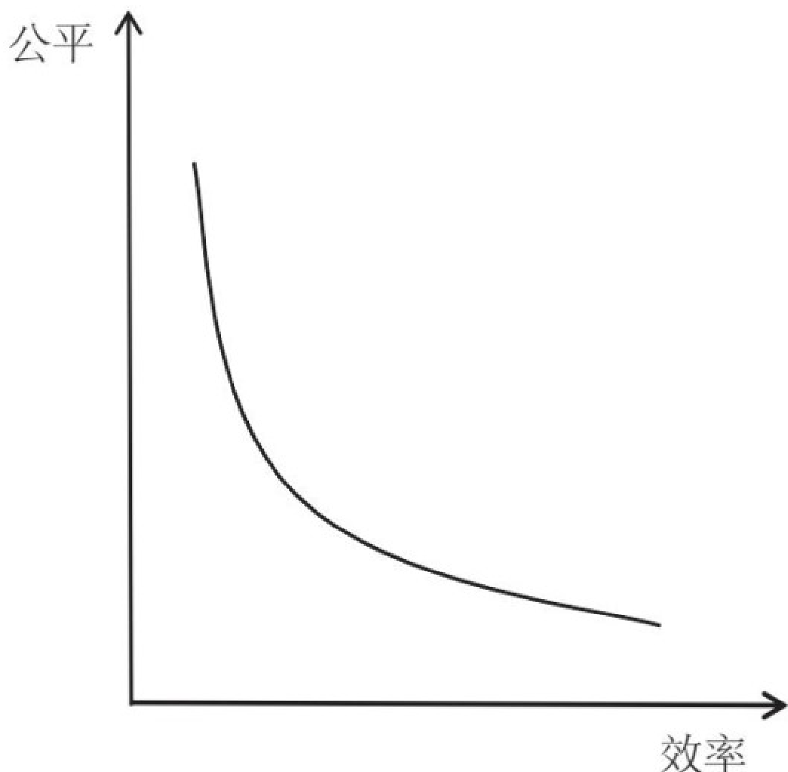


图8–3 效率与公平的悖论

消费公平实现起点公平。之所以把消费公平称为起点公平，是因为在经济循环中改变消费不公平的状态，可以校正经济循环的结果，逐步缩小收入、财产差距，从而为新一轮经济循环的起点——消费，打下更加公平的基础。协同消费带来了消费公平，让更多的人可以通过购买暂时的使用权，减少首次分配时的不平等，使人们能够享有同等的服务，提高每个人的福利水平，改变“没钱不消费”这个逻辑所造成的能力鸿沟。

分配公平带来规则公平。收入分配包括初次分配和再分配两个过程。传统意义上，初次分配注重效率，主要由市场决定；再分配注重公平，主要是政府发挥作用。在共享经济中，分配公平覆盖了初次分配和再分配这两个环节。共享经济使得个人、企业能够相对公平地获得生存和发展所需的人力、财力、物力、机会等生产和生活资料，使得缩小社会成员之间收入差距的可能性增大。实质上，共享经济改变的是在初次分配环节单纯追求效率的规则。共享经济带来的是一种大规模的协作模式，资本、商品、服务的交换逐渐让渡于协作共享，追求个人利益被追

求协同利益取代。更加重要的是，过去由政府主导的公共服务、社会保障、助贫扶弱等领域，也就是再分配环节，将由共享经济带来更多的社会力量和个人力量参与。共享经济在再分配环节所发挥的作用是实现效率与公平的关键所在，也是其与传统经济模式的本质区别。

社会公平成就共同梦想。收入差距过大和分配不公会引发各类社会矛盾，导致社会不安，最终影响经济发展和社会进步。共享经济从一种利润最大化的经济模式转向一种价值最大化的经济模式。在此模式下，公平性大大加强，效率与公平不再是此消彼长的关系，马太效应减缓，社会公平得以促进。美国学者德尼·古莱曾说："发展是达到美好生活的一系列特定的手段。且不论可能的其他目标，发展对一切人群至少具有下列目标：为社会成员提供更多、更好的维持生命的物品；创造或改善与看得到的尊重需要有关的物质生活条件；使人们从自然的、无知的、他者的、体制的、信仰的压迫奴役中解脱出来。"[注]共享经济正在为实现这个目标而努力。

（五）块数据创造经济新动能

美国曾经历过一段由新经济拉动的辉煌，即20世纪90年代的经济高速增长，在科技进步和全球化的基础上，呈现出持续高增长、低通胀、科技进步快、经济效率高、全球配置资源等特点的一种经济状态。有学者把这种状态概括为新经济。2014年，中国提出要顺应全球新技术革命的大趋势，推动以绿色能源环保、互联网等为重要内容的新经济发展，占领未来发展制高点，提升产业和经济竞争力。[注]在不同时空背景下，新经济有了新的内涵。国务院总理李克强说，发展新经济的重点是培育新动能，促进中国经济转型。新经济的覆盖面和内涵很广泛，涉及第一、第二、第三产业，不仅是指"互联网+"、物联网、云计算、电子商务等新兴产业和业态，也包括工业制造中的智能制造、大规模的定制化生产等，还涉及第一产业中的家庭农场、股份合作制，农村第一、第二、第三产业融合发展，等等。而且，发展新经济，小微企业将大有作为，大企业可以有更大作为。[注]

块数据经济带来的是一种新动能，将成为引领新一轮新经济发展的风向标。进入互联网时代，集中化遭遇了碎片化，二八法则被长尾理论战胜，平台经济应运而生，需求集中但客户分散的问题得以解决。块数据经济强调平台集聚，但集聚的不仅仅是需求，还包括创新、运营等各

个方面的价值关联，从而形成新经济的新动能。

从上下游分工转向交互式协作，平台共享带来新动能。基于互联网的协同合作改变了过去以企业为主的产业链上下游分工协作模式，已经对经济生活产生了深远的影响。块数据的运行逻辑在于能够以去中心化的形式配置全球资源，这使得多边市场及全球一体的协作型经济成为可能。比较典型的例子是云经济，其核心特征是交互共享、地域分散及驱动商业价值甚至社会价值。云经济的价值已经在多个领域被不断证明，比如出租车提供按需、按用量付费的运输服务；公司通过“劳动云”购买工人提供的服务。

从分领域创新转向组合式创新，平台创新创造新动能。突破创新壁垒实现创新资源和要素的有效汇聚，是组合式创新的价值所在。21世纪的经济发展格局与20世纪大不相同。创新不再是由大企业自上而下地推进，而是由无数业余爱好者、创业者和专业人士等自下而上地开拓。依托特定平台，调动全社会的创新激情，激活人才、资本、信息、技术等创新要素，形成共同目标、内在动力，进行多方位交流，平台化协作将有效发挥创新对拉动发展的乘数效应。走向“智慧经济”是组合式创新带来的新趋势。智慧经济是一种最优规模经济，它将帮助创业者、企业家、企业组织建立一种具有独立人格、以平台连接、以组合式创新带来新价值的经济模式。智慧经济的竞争优势在于智慧资本，主要是通过品牌资源、概念创新、知识产权、创意发展、智慧的再创新和再消费实现的。

从数量级增长转向指数级增长，平台成长激发新动能。规模思维成就经济增长，但现在需要做的是突破线性思维的局限，从数量级增长向指数级增长转变。这种转变实现的根本，是从过去依靠产品数量的增加转变为客户社群的建立。把客户都集中在平台之上，形成一个价值取向、行为习惯彼此高度认同的群体，分析、研究、发现并满足他们的需求。“商品——人”的关系转换为“人——商品”的关系，大规模定制和生产柔性化得以实现，积水成渊成为可能。比如，3D（三维）打印带来的不仅仅是技术革命，也是一场即将到来的“创客运动”，它把一群爱好者带向了一种全新的经济模式，也就是社群经济。社群经济将成就一大批高质量的小微企业，并把它们整合到创新平台上，这样既完成了平台的裂变，也成就了无数小微企业的几何级成长，由此共同颠覆单打独斗、条块分割的传统模式。阿里巴巴和京东并没有改变商品与人的关系，微

信也没有做到，但社群经济也许可以。新的社群模式和技术进步将催生无数新的可能性。

第二节 区块链与块数据金融

（一）区块链技术

金融是现代经济的核心。互联网金融能否成为下一个蓝海，也许将取决于区块链技术的发展。区块链技术是与比特币同时产生的，比特币的开发者中本聪在2008年年末首次提出区块链的概念。他把区块链定义为“一种通过去中心化和去信任的方式集体维护一个可靠数据库的技术方案”。比特币的出现或者说区块链技术的产生，对整个金融业产生了巨大影响。

区块链的核心技术是去中心化与分布式。目前传统数据库是中心化记录、中心化存储的，即使是异地灾备、云存储，也只是将存储地从一处变为多处，从本地变为云端，如果中央服务器出现问题，则灾备数据库也将停止更新数据。而区块链最大的特征是去中心化，整个网络没有中心化的硬件或者管理机构，全网络的每个节点都依据共识开源协议进行分布式记录、分布式存储、分布式传播，节点与节点之间的权利和义务都是均等的，且任一节点的损坏或者失去都不会影响整个系统的运行。

分布式记录。区块链与传统网络记录体系不同，它没有设立中心记录者，而是全网每一个节点在参与记录的同时也在验证其他节点记录结果的正确性。

分布式存储。区块链是全网全部数据同时存储于系统的所有节点中，只要全网有一个节点保持正常运作，这条主区块链就可以被完整地读取。

分布式传播。区块链中的每一笔新交易都采用分布式的结构进行传播，消息由单个节点直接发送给全网其他所有的节点。

区块链的主要特征是高安全性、高容错性、高包容性。区块链本质上是由交易各方信任机制建立起的一个完美的数学解决方案，具有高安

全性、高容错性、高包容性的特点。高安全性主要是指区块链形成的数据记录不可篡改，任何价值交换活动都可追踪、可查询，这种透明的数据管理体系在法律层面上是无懈可击的。高容错性主要是指区块链的分布式模式使得数据库中的一个或几个节点出错，不会影响数据库的整体运转，更不会影响现有数据的存储与更新。高包容性主要是指基于区块链技术建立起来的数据库足够大，所有的价值交换活动，包括开户、登记、交易、支付、清算等，都可以在这个数据库中完成。

区块链的基本功能是价值转移与价值交换。基于TCP/IP协议（网络通信协议），信息可以快速生成并且复制到全世界任何一个有网络的地方，但始终不能完成价值转移和信用转移。区块链技术使得系统中的参与方无须相互信任就能够完成各种类型的交易和协作，这就为实现价值转移提供了可能。

这里的“价值转移”[注]，是指在网络中每个人都能以可被认可和确认的方式，将某一部分价值精确地从某一个地址转移到另一个地址，而且必须确保价值转移后，原来的地址减少了被转移的部分，而新的地址增加了所转移的价值。这些价值可以是货币资产，也可以是某种实体资产或虚拟资产，价值的转移必须获得参与方的认可且不受任何一方的操纵控制。比特币就是这种价值转移的完美实现手段。

（二）数据信用体系与自信用社会

区块链之所以能够实现价值转移的功能，一个重要前提是它实现了去信任化。区块链依靠非对称加密和可靠数据库，实现了对信任的自行传递。通过算法程序对所有的规则提前进行表述，参与方只要信任共同的算法就可以建立互信，不需要知道交易对手的信用度，更不需要借助第三方机构来进行交易背书或者担保验证。因此，节点之间无须公开身份，系统中每一个参与者的隐私都能受到保护。某种意义上说，区块链建立的是一套数据信用规则，它引发了人们对数据信用体系建设的思考以及对于自信用社会的畅想。

数据信用体系的建立需要重点完成两个方面的工作，一方面是建立一套严密的数据信用规则，实现数据即信用；另一方面是完善信用信息共享机制，实现信用数据化。

区块链技术通过数据区块取代了目前互联网对中心服务器的依赖，

使得所有数据变更或者交易项目都记录在一个云系统之上，理论上实现了数据传输中对数据的自我证明。深远来说，这超越了传统和常规意义上需要依赖中心的信用验证方式，产生了一种“基础协议”[注]。通过这样一套数据信用规则，每一个参与者都能够守信并获得他人信任。

传统的线下社会，数据采集缺乏和信息的快速流通造成了征信不便，而在万物互联和大数据技术发展的基础上，信用的计算和共享成为可能。单一的个人数据记录无法对个人画像，但把离散的数据集聚起来进行关联分析后就会有非常明确的指向。同样，通过对任何企业和组织的关联数据进行分析，也可以得出企业的全息画像。如数联铭品（BBD）作为一家大数据金融服务机构，利用大数据技术追踪数以万计的数据源，通过数据采集、数据清洗以及数据关系分析，能够在短时间内输出完整专业的企业背景信息。

大数据时代最需要信任、信誉和信用。互联网连接几十亿节点和用户，破解“网络另一端与你聊天的可能是一只狗”的困局，快速建立互信成为关键。目前，经济运行中现实的做法是建立供应方、平台方、需求方的信任联盟，保证共享交易的执行。从更长远来看，建立自信用社会体系是一种必然要求。自信用社会建立的前提和基础是数据信用体系的形成，而数据信任体系建设的核心是人人信用公开。在互联网、大数据和物联网技术的支持下，通过建立社会信用征信系统和信用评估体系，实现信用信息的共享和查询，清除信用信息交流的屏障，使每一个市场主体的信用数据都可查可用，防范信用风险。当然，自信用社会的形成也不是完全由数据信用体系决定的，还受到社会风气、经济环境、开放程度、宗教信仰等多重因素的影响。

（三）块数据金融新构想

区块链技术始于比特币，是电子货币层面的规则创新，这样一套规则有可能更好地从电子货币层面向整个金融层面转移，尽管实践推进能进展到什么程度仍未可知，但区块链的确给我们观察即将到来的新金融时代提供了视角。

泛金融化，更加高效。新金融时代，随着互联网技术、大数据和云计算等新一代信息技术的发展，金融信息化并不止于让手机代替柜台那么简单，它正在金融业内和业外引发一场实实在在的革命。金融机构在增强传统金融业务的同时，也在不断地拓展服务的范围和内容，提供

更加综合、更加社会化的服务。客户已经被高度数据化，通过关联数据对客户进行金融画像描述从而精准评估客户的金融需求，实现金融产品的精准推送和金融服务的高度个性化定制。需要更加关注的是，金融供给不再是传统金融机构的专属领域，许多具有大数据技术应用能力的企业都会或多或少涉足、介入金融行业。而金融企业与非金融企业间的跨界融合也将成为常态。

金融脱媒，更加便利。金融脱媒是指在金融管制之下，资金供给绕开金融媒介，直接输送给需求方和融资方，完成资金循环。现在普遍认为P2P（个人对个人）借贷和众筹融资是金融脱媒的表现，但实际上二者只不过是把金融媒介由金融机构转移至网络，事实上在需求对接过程中依然有媒介存在。金融媒介分为两种，一种是货币媒介，一种是交易媒介。货币媒介经历了从无到有的发展，随着科学技术的发展，还必然经历从有到无的过程。数字货币的出现，使物与物之间的交换实现了货币脱媒。交易媒介可以分为组织、平台和合约三种类型。目前通常所说的金融脱媒是指在金融管制之下，金融运行绕开组织中介和合约中介，直接通过平台中介实现。特别是随着数字账户和数字信用的实现，这种金融脱媒的趋势会更加明显。金融脱媒的实质并不在于“脱媒”，而在于提高经济交易活动的便利性和透明度，实现金融权利的觉醒。

普惠金融，更加公平。随着金融监管的放松、金融自由化的推进以及信息技术的发展和应用，人们开始对于金融本质及其与社会发展、人民福祉的关系进行思考。在金融失衡和金融权利不平等的背景下，金融创新展现出了完全不同的新方向和解决方案，比如普惠金融。普惠金融的基本含义是能有效、全方位地为社会所有阶层和群体提供服务的金融体系，目的是使弱势群体和低收入群体能够以合理的价格，方便且有尊严地获取全面、高质量的金融服务[注]；使金融像其他经济服务一样，回归促进价值交换、优化资源配置、托管社会财富的本质。

第三节 痛客：无痛点不创新

（一）新经济崛起与价值重构

任何一个成功的经济模式都有一个共同的特点——满足需求。由于新需求的产生，社会化大生产加快了重心转移，即基于消费升级的物质

产品大生产向精神产品大生产的转移，而这将引发新经济的崛起。

新资源：尚待满足的需求主张。需求是新经济时代经济成长的重要因素，经济效益归根结底取决于市场需求。尚未满足的需求主张在哪里？需求的主体是人，需求的对象是商品与服务，满足需求归根结底解决的是人的发展问题。中国经济正在进行新旧动能的转换，亟待培养新的动能，而民生改善将孕育发展新动能。通过不断保障和改善民生，满足不断升级的多样化需求，就能拓展更大的发展空间。

新路径：尚需激活的创新潜能。创新是打开需求大门的一把“金钥匙”，也是经济社会发展的“强引擎”。中国经济发展的每一次破茧成蝶，靠的都是创新，不单是技术创新，更包括体制机制创新、管理创新和模式创新。[注]创新将提高对需求变化的适应性和灵活性，只有发现、挖掘与激活需求，发挥创新的原动力作用，形成新技术、新产业、新业态，才能发挥创新驱动发展的乘数效应。

新价值：尚在形成的价值关联。新经济将推动整个经济社会出现多维度、系统性的变革，这一变革的核心就是价值重构。信息化、数字化、网络化、智能化的相互融合，信息技术、制造技术、新能源、新材料、生物技术的相互融合，将催生新的经济结构和经济形态，实现价值重构。比如，互联网建立了制造商与消费者的直接交互平台，由于产品的数据化，消费者在使用产品的过程中不断反馈关于产品的相关数据，而制造商无须接触消费者就可以直接获得这种数据，一些看似无关的因素因为数据化成为价值关联的关键。生产供给与消费需求之间形成了一种共同成长、相互依存的关系。传统的价值链被打破，新的价值关联形成，并将由此产生巨大价值。

（二）痛客与创客

无痛不变，无痛不新，无痛不进。这里的“变”指的是变革，这里的“新”指的是创新，这里的“进”指的是社会进步。从价值的角度说，创新带来的是功能升级，痛点代表的是情感诉求，当二者实现高度关联和碰撞激活的时候，价值就得以实现。

在营销学中，痛点是人们在使用产品或服务时抱怨的、不满的、让人感到痛苦的接触点。转化到社会生活中，就是被大多数人反复抱怨过的有待解决的问题或有待实现的愿望。提出痛点的人，即痛客。作为中

国大数据产业策源地的贵州省贵阳市，就正在启动一个全新的痛客计划。

痛点是问题、是需求，它具有三个鲜明的特征。第一，广泛，痛点是指目前尚未被满足而又被广泛渴望的需求，比如，打车难是痛点，于是有了打车软件；资源闲置浪费是痛点，于是有了网上闲置商品交易市场。第二，迫切，痛点无处不在，但发挥作用的是另一个关键变量，就是迫切性，当一个需求变得非常迫切、急待解决的时候，它就成了痛点。第三，复杂，痛点往往是复杂的，痛点的搜集、综合、存储、加工、挖掘和价值创造是一场非常复杂的博弈，需要同时从多个层面入手。

痛客是行动，是模式。痛客能够从各种社会现象和问题中，找到普遍存在的尚待解决的关键问题，形成并提出痛点。无论是政府机构、企业、社会组织，还是自然人，都可以成为痛客，痛客计划就是一场"大家一起发现"的行动。当前存在这样一种社会共性问题，即有办法的人没想法，有想法的人没办法。痛客计划的提出，在有办法和有想法两类人之间搭建起了互动交流的平台，从而激发大众创业、万众创新。

要实现真正的高附加值的创新，痛客在寻找痛点时必须遵循SMART黄金原则：S，即明确（Specifi c），指痛点要清晰、明确，能够被准确地理解；M，即可衡量（Measurable），指痛点能够被量化或质化，可以采用相同的标准进行衡量；A，即可实现（Attainable），指痛点通过努力可以解决，也就是目标不能过低和偏高，偏低了无意义，偏高了实现不了；R，即与其他目标具有相关性（Relevant），指痛点应该与社会生活密切相关，不是核心问题不能作为痛点；T，即有时间计划性（Time-based），指痛点的完成要有时限性，要在规定的时间内加以解决。注



图8–4 痛点的SMART黄金原则

创客一词来源于英文单词"maker"，是指出于兴趣与爱好，努力把各种创意转变为现实的人。创客的特点是"单打独斗"，但这种单打独斗只是摆脱了对生产资料系统的依赖，转而更加强调价值关联。当前创客

发展过程中存在诸多问题，如创新难以落地，创新重复性严重等，原因就是没有实现价值关联。痛客计划的提出使得这种价值关联成为可能，创客可以成为一批“基于社会痛点，系统提出解决方案的人”。痛点是产生创意、催生创新的土壤。只有聚焦社会问题，回应社会需求，创客的创新举措才能更有针对性，创新成果才会更具价值。创客与痛客相互促进、相互转换，创客会因为痛客的出现而进入新的发展阶段，痛客会因为创客的实践而逐步实现价值。创客与痛客之间的边界被打破，创客可能会成为痛客，痛客也可能转变为创客。

高价值必须通过创新获得卓越的功能，但是功能本身并不能创造出价值，只有当强大的诉求被满足的时候，价值才能够实现。如何看待价值、痛点和创新的关系，痛点和创新如何实现价值关联，可以用一个简单的等式来描述：

价值 = 创新 × 痛点（V = I × P）

即价值（Value）等于卓越的创新（Innovation）与强大诉求（Pain-point）两者的乘积。可见，创客与痛客关系紧密，搭建一个从痛点到创新，从痛客到创客的连接平台，从而构建一条完整的创新创业链条，最终形成一种新的经济形态。这种经济形态强调开放、创新、共享，可以集聚众人的智慧、资源、资金，精准发现问题，系统解决问题，实际上就是众包、众创、众筹、众享的相互融合和相互作用。

（三）痛客经济与社会经济模式

新经济时代，不仅需要一套不断更新的系统性理念，更呼唤着一场能够经得起实践检验的结构性变革。痛客经济是一个全新的理念，块数据经济是一套开创性的发展思路，是对未来新经济发展的深刻思考。

痛客经济具有重要的时代价值，蕴含着4个基本内涵。其一是效益统一，痛客经济以价值最大化为目标，追求社会效益和经济效益的统一，强调通过创新来解决社会问题，满足社会需求。其二是信息对称，依托互联网和新媒体，供给方和需求方之间实现了信息直通，避免中间流通环节可能产生的信息壁垒和成本上升，确保创新的及时性和高效性。其三是循环发展，需求催生供给，供给又带动新的需求，在不断的发展过程中逐渐构建完整的产业链和循环的创新体系，形成可持续的发展动能。其四是智能运行，通过块数据技术的运用，更加科学精确地聚

集社会公众和政府的真正需求，实现对需求端的数字化管理，以更为精准的方式驱动创新创业的实现。痛客经济与大众创业、万众创新的时代背景紧密相关，对于重构创新产业链、重塑新经济形态具有重要意义。

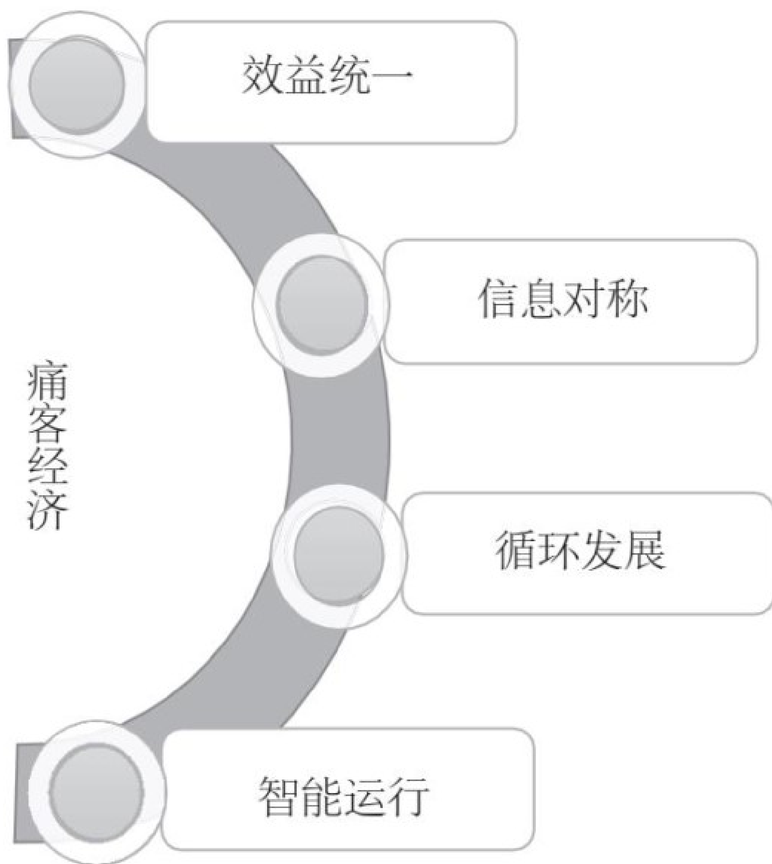


图8–5 痛客经济的内涵

痛客经济的构建并不着眼于一个完整的经济运行体系，它更多的是研究社会经济领域和创新发展领域相互转换的规律和模式。当自然爆炸性增长结束的时候，痛点就关系到产业的兴起和就业机会的创造。让人们获得清洁的水、医疗保障、金融服务、有用的信息，同时解决人们迫切需要解决的一切问题，将成为新经济时代政府、市场、社会共同努力的方向。任何一个痛点只要有人想要认真地去解决，就能创造出成百上千的新的科技构想。其中，最优秀的创新将带来新业态、新模式的发展，并激发更大的新需求和发展的新机会。新经济不仅需要一个科学技术实验室，还需要一个商业模式实验室，需要把思想转换成产品，才能够让$V=I\times P$变为现实。

痛客经济将催生一种新的社会经济模式，产生一大批社会企业。社会企业与政府、企业、非营利组织不同，它没有政府的强制权力和财政收入，发展方式主要靠万众创新；它没有企业的最大化利润和盈余分红，发展目标集中于解决社会问题；它没有非营利组织的外界资助，发展机制更突显自主管理的特点。这样一种新的社会企业与新经济时代的技术创新相结合，将形成一种新的社会经济模式，其优势在于能够实现三个方面的价值。

新的社会经济模式将促进个人价值的实现。社会经济模式是人人时代的产物，它不是一种单纯的经济模式，而是一种社会的共同行动力，以解决社会问题为发展目标，通过发现并挖掘社会中未被满足的需求，用创新的方式高效地整合人力资源和社会资源，使参与其中的每一个人都能够充分实现自我价值，为解决社会问题而贡献力量，成为社会中坚。

新的社会经济模式将突出社会价值的实现。社会经济模式的使命是实现资源的公平分配和社会的和谐治理。资源配置模式与共享经济天然契合，核心都是协同共享，交换价值被共享价值所取代，将有利于各种社会问题的解决，并积累形成社会资本。社会经济模式在理论上和价值观上都将指引人类经济社会走上一条新的发展之路。比如，一些社会企业建立各种医疗、教育等社会服务机构，为有需求的人群提供社会服务。美国有一家美好世界书店，它就是通过售卖二手书的方式，将收益捐给支持扫盲的组织，再由扫盲组织捐给贫穷落后的国家，解决其文盲问题。

新的社会经济模式将带来经济价值的实现。社会企业以企业的方式运营，使其摆脱了传统非营利组织的局限，利用自身的经营创造就业、提供社会服务、维持自身的可持续发展。正是由于社会企业需要以企业的方式运营又要完成社会目标，才会激发出更多创新的管理模式，成功结合社会治理和商业盈利，以高质量的服务和产品服务大众。社会企业的经营管理模式多样，包括合作社、互助协会、社团等形式，同时社会企业并不特指法人结构的企业，而是指利用企业经营模式解决社会问题的组织。

（四）痛客治理与痛客智库

大数据时代，社会治理的难度更高、情况更复杂，需要进一步推动

治理创新。痛点与痛客凝聚的是一种社会共同行动力，不仅会带来经济模式的变革，也将推动治理方式的创新。痛客治理是一种复合治理，表现为价值复合、主体复合、运行复合。

价值复合，实现的是治理理念从“利益”到“价值”的转变。在利益导向的治理理念下，社会风险不断增加，社会矛盾不断累积，建立以价值为导向的复合治理模式是一种必然趋势。痛点与痛客在社会治理中的应用，实际上就是要形成科学、理性、有效的诉求表达机制、利益协调机制、矛盾调处机制、权益保障机制，最大限度地调整社会关系、维护社会秩序，实现善治的社会治理目标。

主体复合，实现的是治理主体从一元到多元的转变。痛点的背后是社会需求的多样化、多层次，痛客引发的是一场“大家一起发现”的广泛的社会行动，实际上就是要改变过去政府大包大揽的管理模式，凝聚社会组织、企业、公民个体等社会力量，让大家的问题由大家来解决，加快形成党委领导、政府主导、社会协同、公众参与、法治保障的社会治理格局。

运行复合，实现自主化、专业化与社会化的结合。当前，治理面临的普遍问题是，谁来发现问题，谁来解决问题，如何建立良性循环的治理机制。痛客发现问题，实现的是治理的自主化。想要实现科学治理，还必须强调遵循治理的客观规律，提高专业化水平，这就需要引入专业的智库机构。在专业化的基础上，进一步推动治理社会化，通过发挥智库的专业作用，把痛客和创客连接起来，把资本和资源整合起来，形成完整的良性循环。

（五）痛客计划

一个地区的经济社会发展，除了需要技术和模式创新，还需要在战略层面进行持续创新。贵阳推出的痛客计划就是这样一个创新与探索。痛客计划是以解决痛点为导向，汇聚以痛客为代表的社会智力资源，并将其与资本要素、价值链驱动等多方面因素进行立体化整合的计划。通过痛客计划，贵阳期待创造一种新的社会需求和解决方案的管理机制，同时为掌握大数据核心技术的人才、拥有大量数据的人和了解大数据核心需求的人提供一个互动交流的平台，进一步深化大众创新、万众创业，助力贵阳打造创新型中心城市。痛客计划分为两个阶段实施。

第一阶段，举办痛客大赛，打造思想孵化器。2016年3月1日，贵阳举行了全球首个以痛客为主角的比赛——首届中国痛客大赛暨社会共治·企业信用痛点主题大赛，首次大规模、系统性地挖掘痛客及痛点的价值。通过对痛点确权，痛点提供者将得到知识产权保护，痛客所提出的痛点的商业价值也将由市场来评定。今后贵阳还将持续营造和倡导痛客文化，提升痛客计划的影响力。

第二阶段，打造痛客梦工厂，搭建产业孵化平台。以痛客大赛为基础，贵阳市将打造痛客梦工厂平台。痛客梦工厂是痛客计划运营的基础平台，即发现痛点的潜在价值，培育痛客的市场价值，实现孵化器自身的增值，发挥连接知识创新源头和高新技术产业的桥梁作用，使其成为汇聚大众智慧成果的重要载体。利用大数据平台的快速处理和高效运转功能，痛客梦工厂将构建一个科学合理的网络状分工体系，实现痛客、创客、投资者、政府各方价值的最大化。

2016年3月1日，中共贵州省委常委、贵阳市委书记陈刚在2016贵阳国际大数据产业博览会痛客计划新闻发布会答记者问环节指出："痛客的提出将会推动贵阳、贵州乃至国家的创新驱动发展。痛客计划使痛点更有效地和创新创业结合到一起，提高社会的效率，是推进经济、社会发展的一个模式，也是在政府引导下推动大众创业、万众创新的一个重大的探索。"

第四节 容错性创新试验

（一）数据驱动性创新试验

新常态，新经济，需要新改革。以数据驱动为核心特征的新经济将激发新需求、新市场、新技术、新业态、新模式的蓬勃发展，它是一个全新的领域，是一个有待探索的事物。要想确定数据驱动性创新将在什么基础上形成突破，将靠谁来完成，需要开展创新试验，需要充分调动积极性和创造性，鼓励创新、包容创新、支持创新。

我国将科技创新作为提高社会生产力和综合国力的战略支撑，实施创新驱动发展战略，强调以科技创新为核心，同步推进制度创新、管理创新、组织创新和模式创新。继批准北京中关村、武汉东湖、上海张江

为国家自主创新示范区后，国家又设立上海、福建、广东、天津四大自由贸易试验区，而后又确定京津冀、上海、广东、安徽、四川、武汉、西安、沈阳8个全面创新改革试验区，以推进系统性、整体性、协同性改革的先行先试。在此基础上，国家发展与改革委员会、工业和信息化部、中央网络安全和信息化领导小组办公室批准了贵州省建设国家大数据（贵州）综合试验区，这是全国首个获批建设的国家级大数据综合试验区，标志着我国数据驱动性创新试验启动。

国家大数据（贵州）综合试验区将通过3～5年时间的探索，打破数据资源壁垒、强化基础设施统筹，打造一批大数据先进产品，培育一批大数据骨干企业，建设一批大数据众创空间，培养一批大数据产业人才，有效推动相关制度创新和技术创新，发掘数据资源价值，提升政府治理能力，推动经济转型升级。大数据综合试验区将围绕八大主要任务开展系统性试验：

一是开展数据资源共享开放试验。大数据综合试验区将制定数据共享开放制度规范，推动政府数据资源交换共享和开放，建设基础信息共享数据库，探索政府数据和社会数据融合创新，建立健全大数据安全保障体系。建成“云上贵州”政府数据共享交换平台和数据开放平台，建成人口、法人单位、自然资源和空间地理、宏观经济等四大基础信息共享数据库，在全国率先实现政府数据共享交换平台和数据开放平台省市县三级全覆盖，依法有序开放公共数据，带动社会、行业、企业及互联网数据开放共享。

二是开展数据中心整合利用试验。统筹政务数据资源和社会数据资源，建设南方数据中心基地，推进数据资源集聚，打造南方宽带网络枢纽，推动数据存储和云计算协同发展。面向本区域、其他区域和中央部门、行业企业等用户，提供应用承载、数据存储、容灾备份等数据中心服务，为建立国家数据中心体系提供探索。

三是开展大数据创新应用试验。在宏观调控、市场监管、社会治理等领域开展政府治理大数据创新应用，实施“数据铁笼”、大数据治税等重点工程，提升政府治理能力。实施“精准扶贫云”示范工程，建立西部贫困地区大数据精准扶贫的示范应用。在健康医疗、交通、旅游、教育、文化、人力社保等领域开展民生服务大数据创新应用。

四是开展大数据产业聚集试验。充分发挥贵州省大数据发展的先天

优势和先行优势，以优势聚资源，以应用带发展，推动传统产业与大数据融合发展，推动大数据在工业、农业和现代服务业的示范应用，发展智能制造、农业大数据、电子商务等新兴产业和业态，带动大众创业、万众创新，建成有特色、可示范的大数据产业发展集聚区。

五是培育大数据产业生态体系，发展大数据存储、采集、加工分析、云平台建设与运营服务业等大数据核心业态，发展智能终端、集成电路、电子材料与元器件、呼叫中心等关联业态，发展电子商务、互联网金融、智慧物流、智慧健康、智能制造、智慧旅游、智慧农业、智慧能源等衍生业态。打造大数据金融服务平台，推进“互联网＋”普惠金融发展。打造一批满足大数据重大应用需求的产品、系统和解决方案。

六是开展大数据资源流通试验。以贵阳大数据交易所为依托，建立大数据资源流通与交易服务平台，培育一批大数据资源流通与交易服务市场主体，丰富大数据资源流通产品体系和交易模式，建立完整的大数据资源流通与交易机制、制度和标准，打造大数据资源流通与交易的生态圈，促进大数据跨行业、跨区域流通和开发利用，建成全国重要的大数据资源流通与交易中心。

七是开展大数据国际合作试验。积极参与大数据相关国际合作框架体系内的国际研发和项目交流，打造“数博会”等国际会展交流平台。引导国内外企业加强大数据关键技术、产品的研发合作，推动我国大数据产品、技术和标准走出去。

八是开展大数据制度创新试验。将服务模式创新、政策制度突破、体制机制探索作为大数据试验区建设的重点，建立大数据地方法规规章，推动数据权益保护立法、个人信息和隐私保护立法、数据安全管理立法，建立关键共性标准，探索建立有利于推动大数据创新发展的政策体系。

（二）数据共享与数据监管

数据驱动性创新试验的一个重要前提是数据的共享与开放。国务院颁布的《促进大数据发展行动纲要》强调要推动各部门、各地区、各行业、各领域的数据资源共享开放。但现实的情况是，数据共享面临掣肘。究其原因，主要是数据监管存在偏差。

从数据监管的角度来看，存在过度监管和监管不足并存的问题，而

过度监管和监管不足都将带来风险。从数据所有者和使用者的角度来看，存在过度保护与无限共享的偏差。这些偏差使得汇聚了大量优质数据资源的各级政府和公共机构，在数据共享开放上出现了“不愿、不敢和不会”的问题，除了部分自用和有限的信息公开外，数据资源作为“生产要素、无形资产和社会资本”的作用基本没有发挥。

不愿共享开放是认识问题。一些政府部门和公共机构尚未意识到数据共享开放的价值。此外，也存在利益分配的问题，有些政府部门和公共机构把自己掌握和获取的数据当作自己利益和权力的一部分，甚至看成是私有财产不愿共享开放，造成不同部门之间甚至同一部门不同机构之间都难以实现数据共享开放。

不敢共享开放是约束问题。由于我国在数据共享开放方面的法律法规、制度标准建设相对落后，没有形成对于数据共享开放的刚性约束，数据共享开放缺乏考核管理体系，数据共享开放价值不明确、市场不健全，相关部门人员担心政务数据共享开放会引起信息安全问题，担心数据泄密和失控，对数据共享开放具有恐惧感，不敢把自己掌握的数据资源进行共享开放。在我国，《保密法》中对定密、解密程序、泄密处罚以及救济机制等重要制度的设计已落后于数据开放的现实情况，严重束缚了数据的共享开放。

不会共享开放是能力问题。政府数据共享开放缺乏专业性，没有进行分级分类，难以实现收放结合。政府数据该共享开放但却没能实现会引发数据隔离与封闭、价值损耗、信息孤岛等一系列问题；相反，不该共享开放而共享开放，或者不该大范围共享开放而大范围共享开放也可能带来更大的风险和隐患，甚至威胁国家安全。目前我国尚未对数据权保护进行立法，数据共享开放原则、数据格式、质量标准、可用性、交互操作性等缺乏标准和规范，严重制约了大数据作为基础性战略资源的开发应用和价值释放。

（三）试错机制、容错机制与纠错机制

中国的改革，尤其是在大数据领域的创新，充满了“奈特式的不确定性”[注]。需要建立试错、容错与纠错机制，本质上就是建立一套规范的激励与保障机制。试错、容错与纠错，并行不悖，缺一不可。2013年，上海市通过了《关于促进改革创新的决定》，鼓励排头兵、先行者担当起改革先锋的责任，提出在法制框架下为试错护航。随后，深圳、天

津、杭州等地也出台了鼓励创新、宽容失败的规章制度，对容错免责做了明确的规定。贵阳在发展大数据、推动科技创新的道路上也提出了容错精神。只有建立一套试错、容错、纠错机制，才能保障大数据创新源源不断地迸发创造力。

试错，就是要允许创新中出现风险和失败。大数据的发展是一个新事物、新现象，必然面临新问题、新挑战，必须有一个允许创新的环境与机制。无论是个人还是企业，无论是政府还是市场，在创新发展中没有模板可以套用，没有成功的经验可以吸取，必然会有一段“摸着石头过河”的经历，只有通过不断试错，才能找到有效的路径和方法。因此，以试错机制保障创新者无罪，保证创新的“领头羊”不致成为“替罪羊”，就显得非常关键。

容错，就是当试错过程中出现了重大损失和决策失误的时候，启动相应的程序，进行及时止损，并对相应的责任人实施豁免。允许试错是正视失败的一种态度，更重要的是，应尽可能规避错误。允许在创新中试错的关键是设立止损值和止损位。同时，通过科学的容错机制，防止有意试错，减少无意犯错，建立基于块数据技术的监督和考核评价机制，让创新在阳光下进行，谨防少数人钻容错的空子。容错的前提是法治，应有试错的免责条款。容错机制不仅是一个防范风险的技术问题，更重要的是一个文化问题，要在全社会形成一种宽容失败、允许试错、有错必改的氛围。

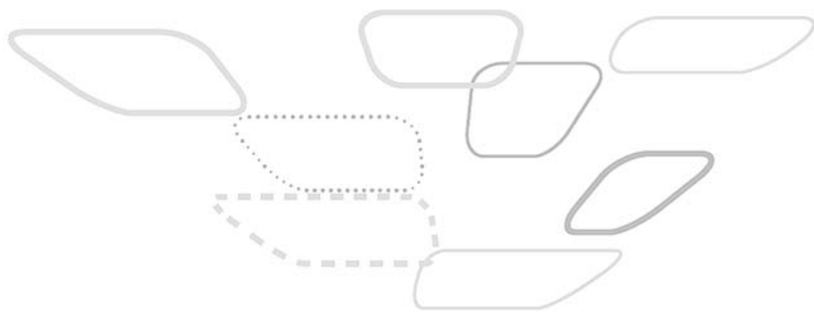
纠错，就是在风险发生之后，有一种机制和程序自动启动，对产生风险的源头、过程及后果进行科学评估，找出原因，修正体制机制，及时引导创新方向，消除路径依赖，防止出现棘轮效应[注]。责任机制是一种重要的纠错机制，避免行为失误的关键就是要建立并落实责任机制。完善的责任机制是改革与创新始终保持正确方向的前提，是始终保持战略定力的基础。

---

1. 熊彼特在《经济发展理论》一书中提出创新理论以后，又相继在《经济周期》和《资本主义、社会主义和民主主义》两书中加以运用和发挥，形成了以创新理论为基础的独特的理论体系。
2. 牟焕森：协同消费的商业模式创新研究，《探求》，2013年第1期。
3. Web2.0，是相对于Web1.0的新的时代，指一个利用Web的平台，由用户主导而生成的内容互联网产品模式。
4. 《零边际成本社会》一书简体中文版已由中信出版社于2014年出版。——编者注

5. 帕累托最优是指资源分配的一种理想状态，是公平与效率的“理想王国”。帕累托改进是指在不减少一方福利时，通过改变现有的资源配置而提高另一方的福利。
6. 德尼·古莱：《残酷的选择：发展理念与伦理价值》，高铦，高戈，译。北京：社会科学文献出版社，2008年。
7. 在博鳌亚洲论坛2014年年会开幕式上，李克强总理做主旨演讲时介绍的新经济主要内容。
8. 全国人大十二届四次会议闭幕后，李克强院总理回答记者问的内容。
9. 梅兰妮·斯万：《区块链新经济蓝图及导读》，韩锋，译。北京：新星出版社，2016年。
10. 梅兰妮·斯万：《区块链新经济蓝图及导读》，韩锋，译。北京：新星出版社，2016年。
11. 联合国把普惠金融定义为能有效、全方位地为社会所有阶层和群体提供服务的金融体系。
12. 国务院总理李克强出席2014年夏季达沃斯论坛开幕式时发表的致辞。
13. 王建武：用SMART原则指导学习型组织共同愿景建设研究，《科教文汇》，2010年第7期。
14. 美国经济学家弗兰克·奈特认为，抛掷一枚熟悉的硬币代表“已知的风险”，概率通常是一半对一半，而抛投一枚未知的硬币则代表“未知的风险”，这才叫作“不确定性”。
15. 棘轮效应：指人的消费习惯形成之后有不可逆性，消费者易于随收入的提高增加消费，但不易于随收入降低而减少消费。

第九章
块数据治理



数据是一个国家的基础性战略资源，同时也是一种宝贵的政府治理资源。从全球范围看，“运用大数据推动经济发展、完善社会治理、提升政府服务和监管能力正成为趋势”。[注]政府是数据最大的生产者和拥有者，数据治理已成为政府治理能力现代化的核心。

数据治理和治理数据，共同构成政府治理的两个维度，打通了未来数据政府的“任督二脉”。数据治理强调政府要实现基于大数据的治理转型，以数据为核心实现政府流程再造；而治理数据则强调政府要做好对数据、数据行为的治理，同时推进数据惠民，增进全社会的数据福祉。

从前者看，我国在政府数据开放、共享和应用方面已开始起步，通过数据推动政府转型、职能转变，包括公共数据的开放在内的努力已经有所进展，展现出巨大的数据优势和发展潜力。但从后者看，对数据的治理还处于空白状态，真正的数据惠民还有待破题。必须看到，今天我们不把数据管好，日后对数据的依赖越深，便越容易出现不可逆的问题，想补课都来不及。

第一节 数据政府

对于政府这个存在，一直以来人们是又爱又恨。本来，政府代表和维护人民的利益，政府官员是人民的公仆，如果人类社会缺少了政府这

一机构，整个国家都将陷入政治秩序崩溃、市场机制失灵和公共事业衰败等一系列风险。伴随着政府的日渐庞大，尤其是它一旦和权力媾和，就会暴露出某种可恶和可怕，包括官员的腐败和政府的俘获，包括强权和独裁的横生等。因此，英国学者约翰·洛克在《政府论》中，就把政府看作“必要的恶”，而明确提出“有限政府”的概念，“宪政”就是要“限政”。马克思和恩格斯则把国家称为社会的“累赘”和“肿瘤”，“最多也不过是无产阶级在争取阶级统治的斗争胜利以后所继承下来的一个祸害”。注但我们能让政府总是以这样一种状态而存在吗？难道政府就只能这样背负恶名而走向未来吗？

不能，也不会！由于政府数字化战略的持续推进、数据资源的大规模开发、数据生态环境的改变等，政府治理将进入一种数据新常态，这将在所有层面和全部领域引发一场政府治理的革命。

数据作为政府的重要资产，同时也是政府治理的重要手段，更是政府治理的成果体现。通过块上集聚形成一种具有内在关联性的数据，预示着广泛的公共需求和公共问题，蕴含着巨大的价值和能量。这些数据深刻地改变着政府的治理理念、治理范式、治理内容和治理手段，将彻底改变传统的以信息控制与垄断来维护威权的治理模式，真正建立起一套“用数据说话、用数据决策、用数据管理和用数据创新”的全新机制，最终把权力关进“笼子”，实现创建法治政府、创新政府、廉洁政府和服务型政府的目标。而建立在这样一种数据新常态下的政府，将不再背负前述种种“恶”名，它将作为一种前所未有的“善”为人们带来全新的治理体验。

（一）政府治理和数据治理

大数据这只蝴蝶的翅膀扇动已经并且正在继续掀起政府治理领域的一场风暴。早在2005年，美国学者威廉·艾格斯就指出，“网络技术正在改变整个政府机构的行为和使命”。注互联网给国家治理带来无限可能。美国、欧盟等发达国家和组织已经相继制定和实施了一系列关于推动大数据国家治理的战略性文件，大力推动政府的数据化转型，特别是围绕更好地发挥数据资源的战略作用，从维护国家安全的高度来增强政府对于未来网络空间和数据主权的保护能力，从而有效提升国家的数据治理力和数据竞争力。毫不夸张地说，“得数据者得天下”已然成为各国政府的普遍共识和一致行动。套用美国管理学家、统计学家爱德华·戴明

的“除了上帝，任何人都必须用数据来说话”，除了上帝，任何政府都必须用数据来治理。

表9–1 大数据战略：全球各国政府和组织机构在行动

国家/组织	时间	行动计划
美国	2009 年	《利用数据的力量服务科学和社会》
		建立全球首个数据开放的门户网站
		《透明和开放政府备忘录》
		《开放政府指令》
	2010 年	《规划数字化的未来》
	2011 年	《联邦政府云计算战略》
	2012 年	《大数据研究和发展计划》
	2013 年	“数据–知识–行动”计划
		《政府信息的默认形式就是开放并且机器可读》
		《支持数据驱动型创新的技术与政策》
	2014 年	《大数据：把握机遇，守护价值》白皮书

国家/组织	时间	行动计划
欧盟	2011 年	《开放数据：创新、增长和透明治理的引擎》
	2012 年	《释放欧洲云计算服务潜力》
		《云计算发展战略及三大关键行动建议》
	2014 年	《数据驱动经济战略》
		《数据价值链战略计划》
英国	2009 年	《政府部门推特使用指南》
	2012 年	《开放数据白皮书》
		“数字化战略”及建立世界上首个“开放数据研究所”
	2013 年	设立全球首个综合运用大数据技术的医药卫生科研机构
		《把握数据带来的机遇：英国数据能力战略》
法国	2011 年	启动“Open Data Proxima Mobie”项目，挖掘公共数据价值
	2013 年	《数字化路线图》
澳大利亚	2010 年	“超级国家宽带网工程”
	2012 年	《澳大利亚公共服务信息与通信技术战略（2012–2015）》
	2013 年	《公共服务大数据战略》
		《数据中心结构最佳实践指南》
日本	2012 年	《面向 2020 年的ICT综合战略》
	2013 年	《创建最尖端IT国家宣言》
新加坡	2006 年	“智慧国 2015 计划”（iN2015）
		“整合政府 2010”（iGov2010）计划
	2011 年	“电子政府 2015”（eGov2015）计划
	2014 年	“智慧国家 2025”计划

国家/组织	时间	行动计划
韩国	2006 年	“U–Korea”发展战略
	2011 年	“智慧首尔 2015”计划
		“构建英特尔综合数据库”
	2012 年	就大数据未来发展环境发布重要战略规划
	2013 年	“政府 3.0 时代”计划
		《第五次国家信息化基本计划（2013–2017）》
		建立“韩国大数据中心”
	2014 年	《2014 信息通信广播技术振兴实施规划》
		未来增长引擎执行计划
联合国	2012 年	“全球脉动”计划
八国集团	2011 年	《开放数据声明》
	2013 年	《G8 开放数据宪章》

但这里要指出的是，对于数据治理的理解，不能停留在目前的数据驱动决策、驱动管理、驱动创新的运作层面，更重要的价值是实现基于数据的治理。这种数据的治理是一种极致扁平、无限开放、高效运行的政府治理。它应该包括但不局限于如下特征：

一是数据治理打破了原有公权力对数据传播流向和内容的控制与垄断，极大地提升了政府治理的“能见度”，通过量化不同事物之间的数理关系，实现一种极致治理。以往，遇到某些重大事件或为解决某一重大问题，政府往往会成立一个由各职能部门组成的工作组，通过信息共享、工作联动的方法加以推进。但是，块数据能够比领导小组做得更好。它能够以更接近于零的边际成本，构建出整个政府和社会数据资源之间的全连接、全流程和全治理框架。这“三个全”打通了政府部门、企事业单位之间的数据壁垒，实现了合作开发和综合利用，有效促进了各级政府数据治理能力的提升。

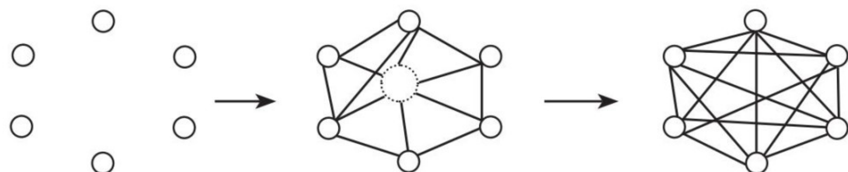


图9–1 数据治理强连接形态演化历程

二是数据治理为公众的直接政治参与提供了平台，政府权力逐渐流向社会。政府和社会在数据共享中实现深度沟通与合作，推进政府治理水平现代化。大数据时代，国家治理强调多元协同，从过去一个主体变为多个主体，从过去由上至下的单向管理变成各个方向协调的治理。“大数据与互联网、微信、微博等新媒体深度融合，可以突破时间和空间的限制，从更深层次、更广领域促进政府与民众之间的互动，形成多元协同治理的新格局。”[注]在全能型政府传统根深蒂固的中国，协同治理要真正落地，块数据应用不可或缺。2014年8月3日，云南省昭通市鲁甸县发生6.5级地震，许多互联网公司利用数据平台和技术优势，纷纷参与到抗震救灾的队伍中，如奇虎360公司开放“灾区寻人”界面，百度公司同步推出“百度救灾地图”。互联网公司将抗震救灾一线的信息需求和大后方的救援资源通过技术平台打通，让需求方、救援者和捐助者都能从中了解最新信息，极大地提高了抗震救灾的救援效率。[注]效率建立在真实且全面的数据之上，但总体来看，这些年我国各级政府提出的许多观念和口号，如建立服务型政府、平台型政府、智慧型政府等，并未有效落地，部分源于没有技术基础，也源于缺乏有效的沟通、协调和数据共享机制。为此，加快建设与大数据发展相适应的数据政府，是实现政府治理能力现代化的必由之路。

三是数据治理以数据科学为基础，以统计软件和数学模型为分析工具。政府处于数据化的环境中，必须强调事中以及事后的监管和服务，提高监管和服务的针对性、有效性。政府的治理活动应以数据为依据，以相关关系捕捉现在和预测未来，从而实现决策数据化和科学化。例如，2013年我国决定对“单独二孩”政策调整时就运用了大数据的力量，

包括使用国家人口管理与决策信息系统（PADIS）进行数据模拟，有关单位将相关数据和备选政策放入集群服务器组里运算了整整七天七夜，最终决定将“单独二孩”政策，由原来的“夫妻双方均是独生子女”放宽为“夫妻一方是独生子女”。需要强调的是，作为数据治理基础的统计数据，必须是真实可靠的，不能有任何“注水”。2013年6月，国家统计局就曾曝光了广东省中山市横栏镇“虚报门”事件，统计数据将仅有22.2亿元的工业总产值虚报为85.1亿元，超过实际约63亿元，几乎是实际值的3倍。而这绝不是个案。

四是追求政府治理精准化。这是块数据理论的最集中体现，即在所有部门和环境推动数据的汇聚整合和关联分析，通过数据的自组织和自激活，实现政府数据治理能力的极大提升，特别是对突发事件的预测和应急响应以及普遍的风险防范。同时，数据治理的实现，反过来会极大地支持和推动政府管理和公共治理方式的变革，为通过块数据实现政府负面清单、权力清单和责任清单的透明化管理提供依据。

要通过数据治理加快政府的数据化转型，推动政府组织结构和形态的优化与再造，实现国家治理体系与能力重构，形成数据政府治理新模式。未来政府的数据治理，既有灵活多样且有效的治理机制，更有基于数据的科学预测、科学决策，乃至延伸到诸如司法领域的数据审判、行政管理领域的数据审批等。畅想未来，政府有可能就是一种机器人，政府的一些部门也许可以成为一个数据终端，通过完全智能化的数据关联提供公共服务，政府部门的精简、权力的公开透明都已经不再是问题，庞大的实体政府被真正的数据政府所替代也许不再是一种“乌托邦”。既然AlphaGo这么厉害，为什么它不能作为一种政府而存在，那样所带来的政府治理理念和社会治理模式进步又将是怎样的一种生动！

（二）公共服务和数据惠民

若干年前，一场来自被称为“蓝色巨人”的IBM公司的脑力大激荡催生了“智慧地球”这一全新概念。此后，智慧城市、智慧政府等一系列智慧概念与大数据一起，成为打开未来之门的钥匙。毫无疑问，数据是改进和完善决策机制的重要材料，其效用更多地将体现为政府和组织的明智决策与判断，以及更有效的服务方式与内容的创新。当然，由于体制机制的束缚，尽管数据日益丰富，但在大多数领域，数据带来的公共服务的便利化还远远不够，更有效用的数据服务产品寥寥无几。所谓的智慧城市、智慧政府更多地还是停留在规划上，还挂在领导办公室的墙壁

上。老百姓还在跑路，政府的数据还在沉睡，这被称为一种数据的悖论。而解决这一切问题的关键，就是要在块数据思维的引领下，在数据空间上重新构建整个公共服务的供给体系。这种基于块数据的公共服务和数据惠民，将推动整个数据应用领域的全面升级，提高全社会的数据福利水平，增进“数据民主”。

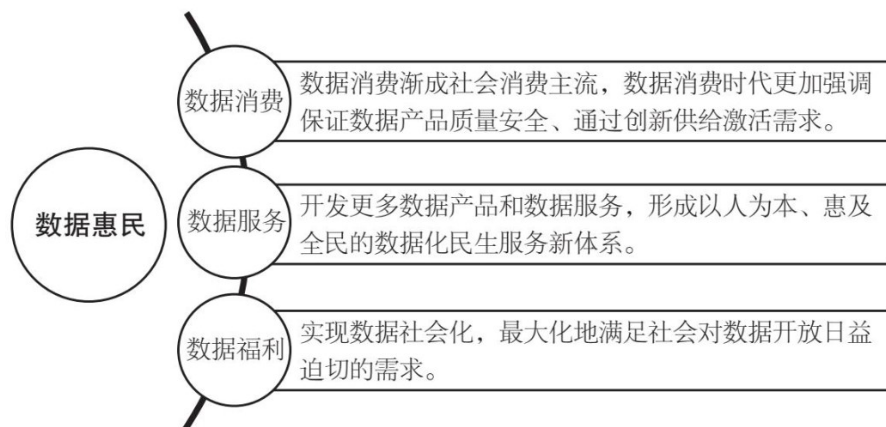


图9–2 块数据与公共服务转变

数据惠民的基础是数据消费，是数据产品的极大丰富和对百姓数据需求的最大化满足。长期以来，我国消费具有明显的模仿型排浪式特征，比如“街上流行红裙子”。但随着大众消费时代的到来，这一消费阶段基本结束，数据消费渐成社会消费主流。用网购的智能体重仪称体重、测体脂，出门用打车软件下单叫车，足不出户用手机订餐……人们的钱包变化开始越来越多地与数据相关，催化了中国消费新变局，一个新的消费时代已经来临。不仅仅是一般的大众数据消费，专业化大数据交易也成为新生业态，如贵阳大数据交易所、上海数据交易中心等相继挂牌运营并完成一系列大数据交易。业内预计，未来的大数据交易将是一个万亿元级别的市场。从国家的层面看，面对数据消费时代的到来，更加强调保证数据产品质量安全、通过创新供给激活需求，开发更多的数据产品和服务，这也成为供给侧结构性改革的重点任务。

数据惠民的重点是政府数据服务能力和水平的提升。围绕建立数据服务的全新范式，通过更多数据产品和数据服务的开发，让政府变得更智慧，让百姓的生活变得更便利，实现更高层次的数据惠民。同时，用数据倒逼政府职能转型，倒逼政府服务升级，倒逼政府流程再造。随着公共服务长尾需求的不断挖掘，还要以定制化的方式开发更多数据惠民产品，满足公众个性化的数据需求。块数据强调关联和聚合，只有高度

关联和集聚的数据才能推动政府公共服务创新和价值再造，形成以人为本、惠及全民的数据化民生服务新体系。未来公共服务数据化的重点，不仅仅体现在公用事业、市政管理和城乡环境方面的数据化，还体现在养老服务、劳动就业、社会保障以及文化教育、交通旅游、质量安全、消费维权和社区服务等领域的全面数据化，依托块数据的集聚，帮助政府更有效地洞察和满足人民群众日益增长的数据需求。

数据惠民将带来整个社会福利的提升，是最重要的民生福祉所在。目前，我国正在加快推进国家政府数据统一开放平台建设，包括建立政府部门和事业单位等公共机构数据资源清单，制定实施政府数据开放共享标准，制订数据开放计划。按照计划，到2018年年底之前，将基本建成国家政府数据统一开放平台。2020年年底之前，逐步实现民生保障服务相关领域的政府数据集向社会开放，实现数据社会化。届时，数据惠民将从战略转变为现实，从文件落实到生活，从而让公众有更多数据“获得感”。

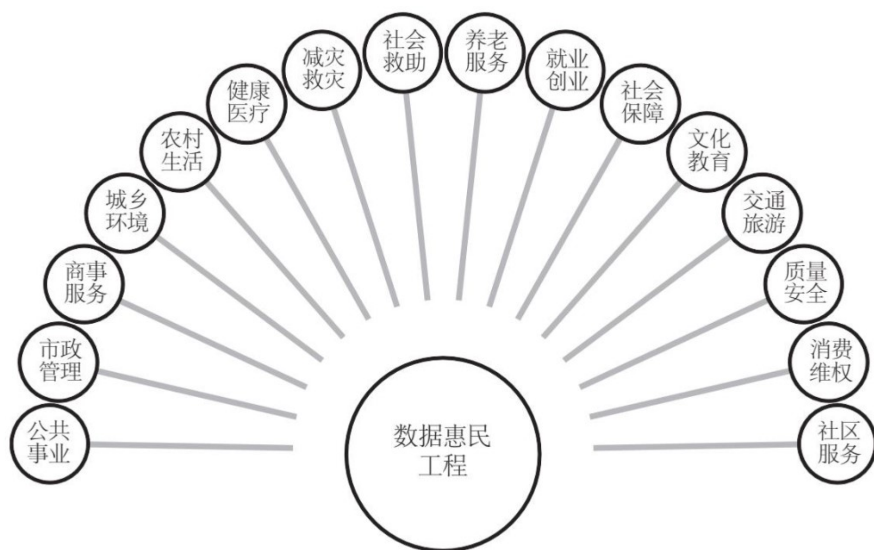


图9–3 块数据全服务链的重点服务领域

（三）廉洁政府和数据反腐

阳光是最好的消毒剂。在治理官员腐败方面，最重要的“阳光”就是数据，数据是最新式的反腐利器。块数据理论强调，权力数据化和数据权力化是未来政府发展的基本趋势和内容，依托大数据实现政府行政流程的再造以及制度安排的优化是廉洁政府建设的要义。

已经有一些地方利用数据技术开展反腐工作，并取得了成果。比如，广东省珠海市纪委已经上线运行了该省首个党风廉政建设主体责任评估系统。该系统运用大数据原理，构建起了一个“责任明确、过程留痕、动态监督、预警提醒、同步督办”的主体责任信息化监督体系。网民通过各种途径关注反腐，他们的搜索、转发、评论等痕迹都可以作为结构化或半结构化的数据被记录。这些数据反映了他们的意见和诉求，成为人们心理变化的记录仪。政府可以利用语义分析和关键词分析，及时掌握公众对反腐的态度，监控反腐舆情，及时回应公众关切。

通过数据公开和透明建立反腐防火墙是数据反腐的重中之重。阿里巴巴经常被批评的一个问题，就是“小二”有很大自由裁量权，他们可能会利用这个权力腐败。阿里巴巴有一个叫“聚划算”的项目，即通过打折、促销，提高用户体验。原来是由“小二”来决定哪个商家能进来、哪个商品能进到这个平台打折促销。在这个过程中，很容易滋生腐败。为此，阿里巴巴数据科学与技术研究院把选择过程自动化、智能化，把这个决定权交给数据，通过大量的数据挖掘、历史数据对比后，自动确定哪个商家、商品应该到聚划算平台上来。这个案例说明，“小二”的权力数据化是杜绝腐败的关键。

数据反腐还有一项十分重要的工作，即建立反腐数据库。最大程度采集和整理各类腐败案例与样本，通过对样本的关键词和关键数据进行提取、标注和关联及聚合分析，编制腐败风险指数，对腐败多发易发领域进行预测及评估，实现块数据层面的预防腐败的全治理链。

第二节 公权治理力

从2 000多年前的烽火台和换马不换人的“八百里加急马上飞”，到1844年美国发明家塞缪尔·莫尔斯向远在64公里外的巴尔的摩城发出世界上第一封电报，再到1876年亚历山大·格雷厄姆·贝尔发明电话。时至今日，超级计算机、云计算平台的发展把人类的数据计算能力推向了新的高度。正如杰里米·里夫金所预言的，第三次工业革命将带来人类权力关系本质的改变和重构。(注)大数据时代的到来，改变了传统公权力运行的轨迹，再造了权力的主体和客体，为我们打开了一扇重新认识权力的窗口。

（一）大数据时代的权力观

对于政府来说，运用数据的特征重塑政府自身模式，进行行政流程再造，核心是用数据对政府组织模式和政府形态进行再造，用数据优化权力运行流程。首先是要树立新的大数据时代的权力观。这种权力观强调，所有治理领域的公权都必须数据化，所有数据化的公权都必须为公众所共享。尤其是涉及公众利益的领域，如公共财政预算收支、公共资源交易和公共项目建设等，都必须实行基于数据化的公共治理，其治理模式、决策过程和审批流程等都可以像扫二维码一样清清楚楚、一览无余。政府的权力清单和权力运行以及如公务员个人财产信息等各方面的广泛公开都将成为惯例，而没有例外。

大数据时代的权力观还认为：数据即权力，权力亦数据。权力是可以数据化的，而数据也将是一种越来越重要的权力，甚至从某种意义上说，谁拥有数据谁就将掌握权力。同时，权力被赋予一种数据的属性，数据分权与制衡成为新常态，权力分散化、权力开放化、权力共享化成为其主要特征。由于权力的数据化，它被处处留痕，它被数据控制，权力的腐败将从根本上被遏制。

“政府之所以为政府，不是因为政府这个名称本身，而是因为与之密不可分的权力的行使与运用。”[注]洛克这句话深刻表明，权力是政治生活的中心，政治的过程就是权力的形成、分配和运用。在大数据时代，权力将随着数据而转移[注]，每一个“数据中心”在一定意义上都将是一种“权利中心”。这些“权利中心”将会拥有话语权，进而推动政府权力向社会的让渡。英国卡梅伦政府已经提出了“数据权”的概念，对于基础公共数据的掌握被视为信息时代每位公民应有的基本权利。大数据把原有官员运用权力的“黑匣子”打开，政府只能以公开和透明的态度回应大数据带来的挑战。2013年5月9日，奥巴马签署名为“政府信息的默认形式就是开放并且机器可读”的行政命令[注]：

> 让信息资源更容易查找、获取和使用，这是开放政府给我们带来的一个重大利好，这些举措能够塑造企业家精神、推进创新、催生新的科学发展……为了继续促进就业，提升政府效率，扩大通过开放政府数据获得的社会利好，新的政府信息默认形式就应该是开放并且机器可读。在整个生命周期之内，我们要把政府信息当作资产来管理，只要可能，只要不违法，我们就要把数据以易于查找、获取和使用的方式发布。

数据治理是政府治理体系和治理能力现代化的基石。就像当年的“德先生”和“赛先生”对中国的影响一样，大数据对政府权力的影响更为直接且广泛。新加坡国立大学郑永年教授甚至还从国家——社会关系的角度写出了《技术赋权》一书，认为互联网提供了一个新的平台，国家和社会在这个平台上竞逐权力，在这样一个舞台上，实现了国家和社会之间的相互赋权。之所以谈到这个问题，是因为科学技术的发展是当代中国努力建立现代民族国家的重要组成部分，或者说是权力合法性的来源之一。数据化的权力是一种新的权力合法化。信息技术的发展创造出一个新的领域和舞台，即包括大数据、互联网等在内的新的公共空间，一旦这个空间被创设出来，政府就无法不受这个新领域的限制。这时，政府和社会就会在这个空间内互动、竞逐权力并展开博弈。在这种博弈的过程中，传统的权力观会发生变化，传统的权力行为模式会发生迁移。通过对权力观念的重塑，政府的权力边界将被深度调整，权力逻辑将被深度改写，新的权力主体开始参与进来，一个更加数据化的权力结构和流程开始形成。

（二）权力的度量

无法计量，就无法管理。块数据主义认为，所有权力都应该而且也能够数据化，所有决策行为都应当摒弃经验与直觉，并无限加大对数据分析的倚重。尤其是随着各种度量手段与工具，特别是人工智能的发展，权力不仅可以数据化，而且还应该是数据智能化。官员的每一次行政行为都应该像下棋一样，每一步都将面对一个AlphaGo，出招是否规范、指令是否合法、决策是否可行，都可以得到人工智能工具的帮助，同时经受人工智能的考量、推演、监督和预警。

对权力进行度量，是块数据应用的一个重要领域，其本质是通过完善大数据监督和技术反腐体系，促进政府简政放权、依法行政。只有权力数据化了，才能装进笼子里，这就衍生出一个概念：“数据铁笼”。现在的问题是，权力的数据化，首先需要对权力进行度量，进行计算。权力是个复杂的社会系统，触角渗透在社会各个领域，不是通过一个简单的四则运算就能进行计算的。据说爱因斯坦说过，在科学上，应该是事情尽可能地简单，直到不能更简单了。权力数据化也应从复杂的大数据入手，以简单的块数据结束。比如，能否找到权力数据化的一般等价物，即更小的、可直接度量的计量单位，让权力度量更简单、更直观。有学者参照基尼系数，提出了权力基尼系数，其最大值为1，最小等于

0。前者表示集体成员之间的权力配置绝对不平均，即100%的权力被一个人全部占有了；而后者则表示集体成员之间的权力配置绝对平均，即人与人之间权力完全平等，没有任何差异。但这两种情况只是在理论上的绝对化形式，在实际生活中一般不会出现。一定社会条件下，社会的权力配置同样存在一个警戒线，即权力基尼系数同样存在一个警戒值，如果大于这一数值容易出现社会动荡。

美国学者斯蒂芬·P·罗宾斯在强调权力关系的特征时提出了依赖性这一维度。他认为："如果你掌握的资源是重要、稀缺且不可替代的，那么人们对你的依赖程度将会增加。"(注)由此，根据依赖性的强弱，就可以对权力的大小进行度量。而根据这种不可代替性，罗宾斯又提出了"权力弹性"(注)这一概念，揭示了决定权力大小的客观因素。

美国学者鸠文内尔提出权力关系的广延性、综合性和强度三 个属性，并引申出计算权力值的函数：$P=\sum_{i=1}^{n} f(c, r)$。这一函 数可以帮助我们从总体上判断某一权力节点上有权者的权力绝对 值，也可以在此基础上通过比较不同权力节点上的有权者的权力 测量其中某一方的相对值。

基于以上观点，块数据理论认为，权力的数据化度量重点是让"大"权力转为"块"权力。同样用一个公式来表示：

$$P=D\cdot M$$

其中，P为权力（Power），D为支配度（Dominance），M为管理量（Span of Management）(注)

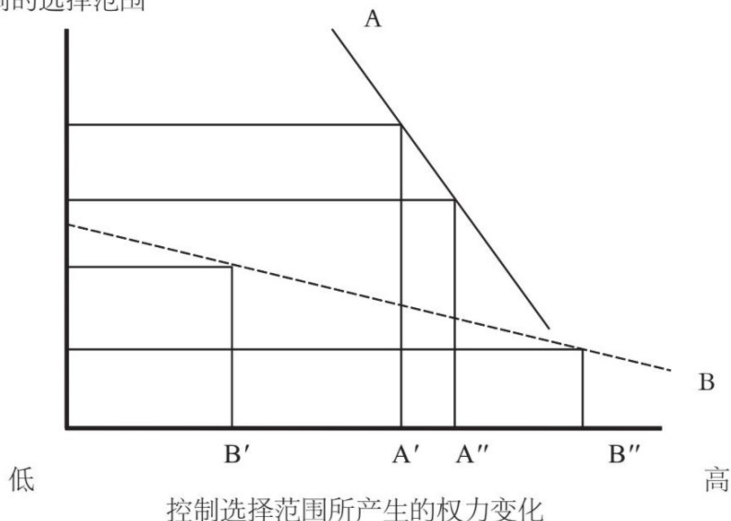


图9–4 权力的弹性

“万物皆互联，无处不计算。”权力的度量是建立健全惩治和预防腐败体系的重要基础。这种度量要紧紧围绕资源、权力、决策等实时数据，借助大数据技术与方法对主客体表征出的大容量、动态实时、多样化、准确的数据流进行分析处理。同时，通过计量，标示出权力的数据刻度和腐败系数，从而加大腐败的成本。

无形的权力应该被有形化、数据化。当传统的权力结构体系被数据深度打破，权力边界被数据深度调整，权力逻辑被数据深度改写，权力才能被关进制度的笼子。大数据的出现是信息平民化的顶峰，它让信息从一种稀缺资源变得人人可以占有、生产和传播，提高了公众的“社会能见度”，并给传统权力带来了前所未有的冲击，从宏观和微观层面解构着权力，实现了权力的回归。所有这些改变，最终转化为一种国家治理的能见度和正能量，为实现真正的“数据治国”奠定基础。

（三）公权治理“四部曲”

权力数据化的核心，是将数据作为思考问题的出发点和落脚点，从顶层设计入手，统一数据标准，提供数据接口，借助新的技术手段，不断提升数据的结构化水平和数据汇聚程度，实现数据治理流程的自动化和智能化，实现“人在干、云在算、天在看”。

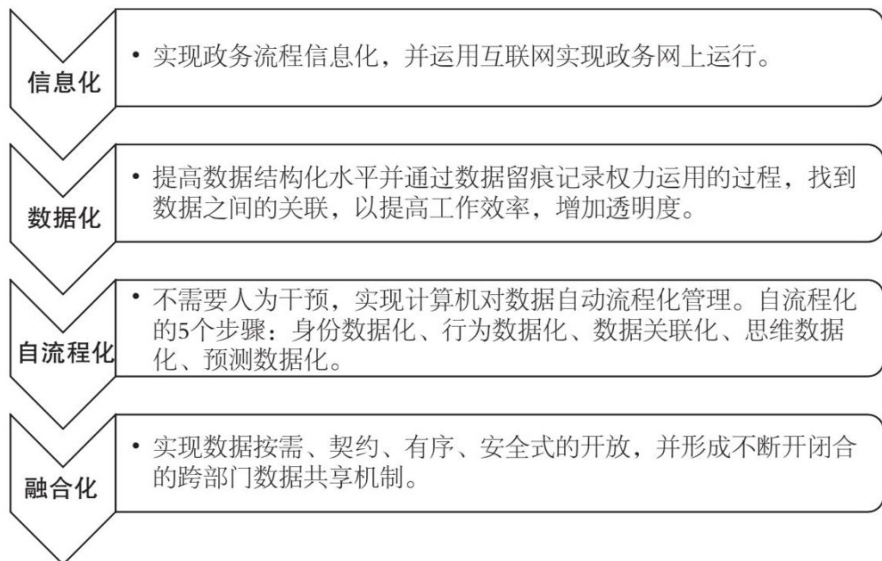


图9–5 公权治理“四部曲”

信息化——夯实基础设施。信息化是块数据治理的基础。通过无纸化、网络化、虚拟化的新方式，借助互联网、云计算和大数据等新兴技术，让政务服务全过程都能够在内外网上流转起来，实现政务流程信息化。块数据治理中的信息化不同于以往“两网一站四库十二金”[注]电子政务工程，更不是简单地将政府行政过程由物理空间转移到网络空间，而是为政务流程中的数据再造提供基础和平台。因此，从一开始就要高度关注数据的来源、安全、汇集等问题，预先做出技术和制度安排，为下一阶段工作打好基础。

数据化——强化数据关联。数据的价值不在于数据有多大，而在于其关联度有多高。块数据治理强调提高数据结构化水平并通过数据留痕记录权力运行过程，找到数据之间的关联，增加透明度。要在信息化基础上进一步实现数据留痕、数据汇集、数据关联、数据分析和数据智能，推动政府的智能化、开放化和公共服务的推送化、个性化。

自流程化——实现流程自动。政府数据治理能力强弱的标准，在于块数据的自流程化管理。面对海量的结构化和非结构化数据，已经无法单纯靠人去分析、去研判。这时，块数据的自动激活、自动预警、自动推送信息就成为关键。就目前来看，可以纳入自流程化的数据对象主要有5类：一是身份数据，即将人和组织进行数据化，建立廉政风险人档案，实现对权力主体身份的自动识别和确认；二是行为数据，把人

和组织的各种行为进行数据化，发现与人的身份数据相关的行为轨迹数据，建立权力痕迹全记录和施政行为数据库，进而把握行为规律和行为缘由，进行行为数据的分级预警；三是关联数据，即由身份和行为数据聚合而产生的数据，反映和发现人与事、事与物之间的关联关系，建立权力风险排查机制和廉政风险预警机制；四是思维数据，即上述数据主体的思维化表达和记录数据，以此分析动机、目的和深层次心理反应，实现风险动机识别和风险来源追溯；五是预测数据，尤其是能模拟和预测风险点和薄弱环节的数据，并对数据进行激活，进而提出廉政风险防范解决方案。

融合化——加快跨界融合。块数据思维的特征在于跨界、共享与融合。利用块数据理念提升政府治理能力的关键，就是要倒逼和打破数据孤岛，一方面引进外部数据，实现外部数据与内部数据相互融合，从而产生新的激活数据；另一方面，建立跨层级、跨区域、跨行业、跨部门数据共享机制，形成数据群岛，最终实现政府和社会的数据相互流通共享、流畅运用和跨界融合。

《礼记·礼运篇》讲大同社会，强调“大道之行也，天下为公。选贤与能，讲信修睦”，化用在块数据理论中，强调的就是“公”是数据的根本，是“大道”，要围绕数据的“公”道，建立数据信用，实现数据睦邻。而“不独亲其亲，不独子其子”，从块数据理论看，就是要打破数据垄断，消除数据孤岛。而“货恶其弃于地也，不必藏于己；力恶其不出于身也，不必为己”，不就是说要推动数据开放，实现数据的跨部门、跨行业和跨领域共享，就是数据人假设中的充分的利他主义。因此，信息化、数据化、自流程化、融合化这“四化”，最终归结到一点，就是建立数据共享社会，达到大数据时代数据治理的最高境界，即数据“大同”。

第三节 “数据铁笼”

“用权的人进了笼子，是因为权力出了笼子。”自权力产生以来，人们就试图控制它，以使它对社会的支配起积极作用。这种控制既需要伦理和道德层面的制约，也需要来自社会和公众的监督，更需要法律和制度“笼子”的规制安排。时任美国总统的乔治·布什于2002年在一次演讲中说：“人类千万年的历史，最为珍贵的不是令人炫目的科技，不是浩瀚的大师们的经典著作，不是政客们天花乱坠的演讲，而是实现了对统治者

的驯服，实现了把他们关在笼子里的梦想。”[注]这段文字被认为是当今政治家对“权力腐败论”做出的最形象的诠释，体现了权力制衡的概念。对于权力的内容要有严肃的规定，对于权力的运行要有严密的监督，对于权力的越界要有严厉的惩戒。

制度，是节制人们行为的尺度或标准，是全体社会成员必须共同遵守的规程与准则。只有不断强化权力运行制约和监督体系，才能形成用制度管权管事管人，让社会公众监督权力，让权力在阳光下运行的价值取向，这是把权力关进制度笼子的根本之策。自柏拉图以来许多人都在问：“谁来统治?”但卡尔·波普尔却认为真正重要的问题是：“如何设计出一种政治制度，令无德或无能的领袖不会对社会带来太大的破坏？”公权力和每个人一样都曾被认为是理性经济人，拥有最大化自己利益的原始动机。任何掌握公权力的人都部分地怀有为自身利益而滥用权力的动机。所以，英国哲学家大卫·休谟才会说，我们应该设计出一系列政府制度，即使流氓占据政府职位时，也将为我们的利益服务。目前，正在贵阳市40多个政府部门全面深入实践的基于权力数据化的“数据铁笼”，正是这“一系列政府制度”的最新形态与综合体现，对个体的描述从来没有这么全面和深刻，对权力的制约从来没有这么科学和具体。这种权力本质的回归之路，体现了一种政治责任和担当，标志着权力制衡与驯服的政治文明新高度。

（一）“数据铁笼”的本质

“数据铁笼”是以权力运行和权力制约的信息化、数据化、自流程化和融合化为核心的自组织系统工程。建构“数据铁笼”的重中之重是回答并解决好“问题在哪里、数据在哪里、办法在哪里”这三个问题，由此优化、细化和固化权力运行流程，确保权力不缺位、不越位、不错位，实现反腐工作从事后惩戒、事中防治转变为事前免疫。



图9–6 “数据铁笼”的三个问题

在本质上，“数据铁笼”强调以块数据技术为基础，实现权力流程数据化、权力数据融合化和权力数据监察化，通过全程采集记录行政行为数据，全面监控行政执法过程风险，精编天网之“经”，密织天网之“纬”，塑造天网之“魂”。建立具备超融合、分布式计算、多维度管理等功能的块数据应用平台，实现对各类权力运行的数据集聚，将分散独立的条数据融合成更具挖掘价值的块数据，构建起预防腐败的电子“防火墙”，确保实现“人在干、云在算、天在看”。建立不敢腐的惩戒机制、不易腐的保障机制、不能腐的防范机制，使不敢腐、不能腐、不易腐成为一种常态。

（二）“数据铁笼”的架构和模式

开放共享的治理理念。“数据铁笼”的核心在于治理，关键在于理念。以先进治理理念引领治理方式变革，以治理方式转变推动治理能力提升，是“数据铁笼”工程建设的重要切入点。“数据铁笼”工程首要的是推动政府数据的开放共享。从权力到流程，从政策到制度，是否可公开、可复用、可问责是开放共享的重中之重。

规范透明的权力体系。“数据铁笼”制约的不是一项权力，而是一个权力体系。这个权力体系包括科学确权、依法授权、廉洁用权、精准

管权、多元督权的全过程。通过权力数据化和自流程化，实时动态监督权力的运行，使每一项权力的运行过程变得规范透明、可量化、可分析、可防控。

跨界融合的平台支撑。以“三统一、四融合”为核心，建立统一平台、统一标准、统一管理，跨层级整合、跨部门整合、跨行业整合、跨区域整合的平台支撑体系，实现关键设施、核心技术、管理体系的统一安全。

持续改进的流程再造。按照“简化程序、减少环节、清单透明、便捷高效”的原则，推动改进政府管理和公共治理方式，推动行政管理流程优化再造，推动政府简政放权、依法行政，实现行政决策科学化、商事服务便捷化和民生服务普惠化。

精准有效的风险控制。强化权力轨迹数据的归集、发掘及关联分析，权力风险预警、风险处置的数据支持、风险预测研判以及智能防控，实现找风险、可评估、能预警、易处置、会防范，真正做到用好权、管住权，从而提高风险控制的精准化和有效化。

多元治理的制度保障。数据“笼子”要靠制度“笼子”来保障，二者不可偏废。建立健全数据“笼子”和制度“笼子”，全面总结“数据铁笼”工程建设中的经验和政策，并使之规范化和制度化。构建以多元主体共同参与为导向，形成党委主导、政府统筹、部门负责、社会参与、法治保障的大数据制度保障体系。

（三）“数据铁笼”的保障机制

“数据铁笼”是一项系统工程，除了要做好顶层战略设计，还要用顶层思维保障落实，明确要求、强化责任，确保项目实施进度和质量。

表9–2 “数据铁笼”的保障机制

保障机制	主要内容
顶层设计	用顶层设计统揽各领域、各行业、各部门“数据铁笼”工程建设方案，用顶层思维指导和推进“数据铁笼”工程建设行动的落实。
基础平台	基础平台是“数据铁笼”工程建设的前提与条件，要超前谋划、统一标准、分级管理、分类建设。
专家指导	组建跨学科、跨行业、跨区域的权威性专家委员会，加强专家咨询、政策研究、业务指导和绩效评估，为“数据铁笼”工程建设提供可持续的决策咨询和智力支持。
工作创新	构建跨部门的政府数据统一共享平台，明确各部门数据共享的范围边界和使用方式，推动政府重要领域和民生保障相关领域数据集向社会开放，加快政府数据资源的融合和关联应用。
安保体系	建立大数据安全标准体系和评估体系，健全大数据环境下防攻击、防泄漏、防窃取监测预警系统，建立完善的网络安全和信息安全保密防护体系，加快数据立法，建立隐私和个人信息管理等保护机制，加强对数据滥用和侵犯个人隐私等行为的管理与惩戒，提高网络与大数据态势感知能力、事件识别能力、安全防护能力、风险控制能力和应急处置能力。
绩效评估	注重定向调控，避免重复建设、资金浪费和资源浪费。强化绩效评估，创新工作机制，充分发挥第三方评估的作用。推动任务考核与工作督查相结合、组织评估与第三方评估相结合、结果评估和过程评估相结合，不断提高工作效率，提升应用效能。

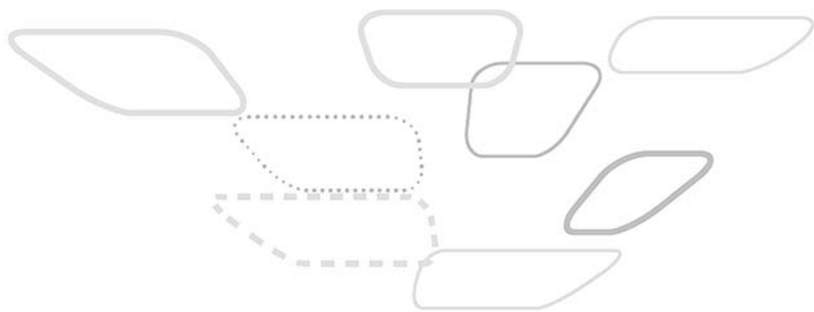
“信任不能代替监督”，在推进建设“数据铁笼”的过程中，要将制度、监督与情感的边界划定得清清楚楚。对于不同性质的权力，可以有不同的制度“笼子”，但最重要的是要有相同的“笼子”，即行政程序法治化。防止滥权或懒权的根本保障在于程序，这也是程序正义的体现。对行政权力而言，具体目标可能因行政领域和地方领域的特性而多种多样，难以合一，但通过“数据铁笼”的顶层设计，可以使这种过程以及程序、原则和制度实现统一。在具体的实施和推进过程中，要注意创新监测方式，运用现代信息技术建立大数据库，有效整合互联网数据和统计数据，全面真实地反映工作落实和实施情况，为科学评估提供支撑。同时，随着相关工作的开展，在接下来，还要积极探索和创新评估方式，引入第三方评估，强化新型智库机构的参与和决策咨询，增强实施的客观性、准确性与可持续性。制度保障程序，程序体现制度，最终使法治

成为程序之治。

---

1. 国务院：《促进大数据发展行动纲要》，中华人民共和国中央人民政府门户网站，2015年9月5日。
2. 马克思、恩格斯：《马克思恩格斯选集》（第2卷），北京：人民出版社，1995年。
3. William D. Eggers. Government 2.0: Using Technology to Improve Education，Cut Red Tape，Reduce Gridlock，and Enhance Democracy. Rowman & Littlefield Publishers. 2005.
4. 贺宝成：大数据与国家治理，《光明日报》，2014年3月27日。
5. 佚名：从百度技术救灾看企业践行社会责任的新模式，赛迪网，2014年8月。
6. 杰里米·里夫金：《第三次工业革命：新经济模式如何改变世界》，张体伟，孙豫宁译。北京：中信出版社，2012年。
7. 哈罗德·D.拉斯韦尔、亚伯拉罕·卡普兰：《权力与社会：一项政治学研究的框架》，王菲易，译。上海：上海人民出版社，2012年。
8. 1990年，托夫勒在他的《权力的转移》中指出，权力作为一种支配他人的力量，自古以来就通过暴力、财富和知识这三条途径来实现。在第三次浪潮文明中，知识将成为权力的象征，谁拥有知识，谁就拥有权力。但知识和暴力、财富不同，后二者具有排他性，一种暴力或财富若为一个人或一个集团所拥有，其他人或集团就不能同时拥有这个暴力和财富；而知识没有排他性，同一种知识可以同时为不同的人所占有。因此，“知识是最民主的权力之源”。谁掌握了知识的控制权和传播权，谁就拥有了权力的主动权。
9. 详情请参阅美国白宫网站，http: //www.whitehouse.gov。
10. 斯蒂芬·P·罗宾斯：《组织行为学》，孙健敏等译。北京：中国人民大学出版社，1997年。
11. 权力弹性曲线显示了A、B两种情形下的权力弹性。A曲线的弹性小，表明A拥有资源的不可替代性强，其相对方通过资源替代而增长的权力小（从A'到A"），相反地，A所拥有的权力就较大。B曲线的弹性大，表明B拥有资源的不可替代性弱，其相对方通过资源替代而增加的权力就较大（从B'到B"），相反地，B所拥有的权力就小。
12. 支配度就是权力主体的自由裁量权，权力可以分为强制约、弱制约、无制约各种板块，把用权方式分解为督导权、终决权、建议权等各种类别，大致估算出支配度。管理量，如一个官员管5万人，另一个管50人，就有量的不同；一个官员支配500亿元的预算，另一个官员只经手50万元，也有量的不同。
13. 两网，是指政务内网和政务外网；一站，是指政府门户网站；四库，即建立人口、法人单位、空间地理和自然资源、宏观经济等四个基础数据库；“十二金”，则是要重点推进办公业务资源等十二个业务系统。
14. 王长华：《司法中立与权力边界》，共识网，2014年9月23日。

第十章
块数据安全



数据的开放共享是把双刃剑，在数据采集、传输、存储、利用和开放中，数据风险问题迫在眉睫，尤其是风险意识和安全意识薄弱、关键信息基础设施的安全可靠性差、黑客攻击、技术环节薄弱和管理漏洞以及法律的缺失和滞后加剧了风险的发生频率和危害程度。

大数据所引发的数据安全问题，并不在于技术本身，而是在于因数据资源的开放、流通和应用而导致的各类风险和种种危机。这种风险和危机的核心是颠覆，颠覆的本质是破坏、解构和重构，直接导致结构和功能的变化，这会加剧社会的不确定性、不可预知性和不可控性。

导致数据风险发生的原因是多方面的，但更为深刻的因素是伦理方面的原因。数据高风险的背后是人性失落、道德失范和行为失序。强化数据安全管理，亟待加强数据安全立法等外部约束性机制的建设。同时，也要重视伦理对个人与社会的影响，逐步建立个人、群体、组织和国家共同遵守的数据价值观与行为准则。数据安全管理需要一个更加系统的框架，从技术、法制、伦理等多个层面加以解决。

第一节 开放数据与数据安全

（一）数据的透明化

数据孤岛和数据垄断是造成数据无法开放的两个重要原因，数据孤岛导致了数据的彼此割裂、互不融通，数据垄断导致了数据所有者对数据的排他性占有。数据孤岛和数据垄断形成的关键因素是数据边界，数据的透明化就是要打破数据边界，使数据开放成为可能。

数据孤岛现象大量存在。现代社会生产生活所形成的数据，多以领域、行业为单位。无论是网络社交平台、各大搜索引擎以及电子商务等形成的企业数据，天气预报、天文观测等形成的科学数据，还是政府各部门形成的行业数据，都存在各自孤立、互不关联、应用上相互脱节的问题。

形成数据孤岛的原因有很多。一方面，数据的标准和接口不统一，各部门、各层级间数据采集重复、标准不同、一致性差、开发利用程度低，导致其在功能上互不关联；另一方面，数据互不交换的处境，会影响数据的记录方式，产生大量的数据冗余与数据垃圾，使得数据、业务、应用互相贯通的模式无法实现。此外，系统间缺乏融合的载体，最终形成数据孤岛。

数据孤岛影响开放共享成本和数据价值体现，严重制约经济社会的发展，以及政府高效透明政务体系的建设。首先，开放共享成本大。各部门、各行业都有自身的数据系统，跨部门、跨行业的数据合作会造成大量的人力物力资源消耗。其次，数据价值难以体现。各部门、各层级间缺乏沟通关联，很难从中发现新知识、创造新价值、提升新能力，更难发现海量数据聚合背后的巨大价值。这不仅会使既有数据的价值得不到充分利用，而且还要浪费其他资源来对其进行管理。

数据垄断也是造成数据无法开放的另一个重要原因，数据垄断的形成与数据的控制力有关。数据的力量源自对数据的控制力。数据权利通常从无权者流向有权者。数字化技术加深了数据贫民和富民之间存在的鸿沟，进一步增强了权利的倾斜。在现实中，诸如政府或龙头企业等数据垄断主体可能会利用数据权利差来获取数据优势。因此导致的可能不是数据或利益的互惠共生，而是数据权利源源不断地从无权者流向有权者。作为大数据时代的产物，数据权利差异不仅会使数据权利者加强对等级森严的机构和社会的控制，同时还会使他们不断追逐对自身权利的支持，从而进一步加深数据权力分配不平等的现状。

数据权利分配差异带来的最大问题就是数据垄断。数据垄断的主体

为了维护自己的利益而拒绝公开共享数据，借口商业机密进行数据保护，加剧数据垄断。各类数据主体都在竭尽全力占有数据、争夺数据和控制数据，并将其贴上自有财产的标签。不能将数据垄断简单地看作一个数量上的差异问题，日益扩大的数字鸿沟带来的数据贫富差异将直接关系整个经济社会的均衡发展，以及社会的公平正义和公共福祉。

政府数据垄断造成公共数据浪费。政府部门对于公共数据的垄断加重了企业负担。通常，政府部门及其相关垄断机构掌握的丰富数据资源正是企业所急需的数据资源，但企业却很难获得这些公共数据。例如，由于政府所掌握的地理位置数据不开放，像百度地图、高德地图这样的企业就得花费更多的成本采集数据进行标注，不仅资源浪费严重，而且效率十分低下。因此，政府数据中的公共数据应当作为一种公共财产归全民所有，既然取之于民，就要用之于民、服务于民，服务于社会创新发展。

商业数据垄断阻碍了数据资源的自由流通。数据孤岛的形成在很大程度上也是由于数据的垄断。从目前的发展趋势来看，通信、银行以及新兴互联网企业凭借着数据采集、分析、利用等技术优势，正大量占有并垄断数据。百度、腾讯、阿里巴巴等龙头企业在呼吁分享开放数据的同时也将用户“圈养”起来，并以这种方式进行数据垄断。这不仅损害了数据资源流通环节，还会影响用户的体验和对于产品的选择。其他领域也存在类似情况——以“开放”为幌子捆绑用户，并在此基础上向其他领域扩张。

数据无边界为打破数据孤岛与数据垄断提供了一种思路。边界在传统企业里可分为垂直边界、水平边界、外部边界、地理边界。1981年，为改变企业的管理层级并消除其中存在的官僚主义，杰克·韦尔奇提出了“无边界”的概念。通过打破企业的4种边界，彻底治愈“大企业病”，建立更快、更柔、更具整体性和创新性的无边界企业。

当前，传统企业的“无边界”概念已逐渐延伸到数据“无边界”上来。数据孤岛和数据垄断带来了数据的边界，企业、政府等数据主体之间为了能实现数据流通，需要打破过去分散在各个业务单元的信息系统边界，通过融合、相互关联集成在一起，实现数据的无边界使用，让数据之间畅通无阻，以谋求更大的发展。

数据的透明化就是要实现数据的无边界化。数据的透明化能够推动

社会发展，最直接的就是带来更迅速、更公开、更灵敏的响应和交流。大数据时代，越来越多的数据被采集，但是数据的种类繁多，不同的组织、企业、部门获得的数据是不一样的，为了能够实现数据的开放共享，就必须让数据透明化、无边界化，这样才能使数据发挥出真正的价值。

数据无边界化是破除数据孤岛及数据垄断的解决方案。只有实现数据开放共享，让公众、企业等机构可以接触到数据，才能做出基于数据处理的新决策，并解决复杂的问题。数据无边界化，就是要实现数据的开放、共享和交易，实现大数据基础设施的开放，实现价值提取能力的开放。让数据真正流动起来，才能释放大数据的价值。

（二）开放数据与政府数据开放

政府信息公开在狭义上专指政务公开，要求行政机关公开其行政事务，属于政府制度层面的公开。而广义的政府信息公开则是在政务公开的基础上，还包括公开其所掌握的其他信息。注

政府信息公开是政府职能的体现。政府是个人、社会和企业之间的衔接桥梁，在各组织之间具有一定的公信力。信息化时代，政府掌握着超过80%的数据，政府采集、管理、占有政务信息，但更重要的是要公开信息。政府信息公开可以促进依法行政，而且对于企业和个人而言，政府公开信息可以有效保障公众获取信息的权利。政府主动公开信息并推动信息共享能够保证民众主动参与的权利，提高公共政策制定的透明度，提高基层对政府政策的响应速度，有效提升政府的公信力。注

瑞典是世界上第一个建立政府信息公开法律制度的国家。1766年，瑞典制定的《出版自由法》成为世界上首部信息公开法案。该法规定了官方文件公开的原则，赋予了普通市民要求法院和行政机关公开有关公文的权利。1951年，芬兰颁布《公文书公开法》。挪威和丹麦在1970年、荷兰在1978年相继公布了有关信息公开的法律。20世纪中叶，但凡不涉及国家安全、个人隐私的美国政府的会议记录（包括国会议员们的讨论记录），都会在网上公开。这一行为被认定为最早的信息公开，美国也成为世界上政府信息公开程度最高的国家。

中国的政务公开始于1985年，在探索农村家庭联产承包责任制的过程中实行村务公开。2008年，《中华人民共和国政府信息公开条例》的

实施，明确了政府信息公开的范围、方式和程序，以及监督和保障的相关措施，标志着我国政府信息公开进入制度化、规范化、流程化的快速发展阶段。2016年年初，国务院办公厅印发了《关于全面推进政务公开工作的意见》，进一步推动我国政务公开的发展。纵观我国政府信息公开的全历程，政府网站已成为其主要平台，政府信息公开工作年报逐步迈向常态化，政府社会治理和社会服务已然成为民众的聚焦点，政府信息公开的力度逐年加大。

但是，政府信息公开不等同于政府数据开放。首先，二者的内涵不同。政府信息公开主要停留在政府法规、流程、权力等方面，更多的是规章、制度等信息层面。真正意义上的政府数据开放主要是指原始数据的开放。2012年德国内政部发布的《德国数据开放》报告中，将政府数据开放定义为公共行政机构所有的能被第三方重新利用的数据。[注]同年，英国政府《开放数据白皮书》将开放政府数据定义为公共领域的信息以及开放给公众使用的数据。美国奥巴马政府在其政府数据开放的八大原则中第二条就指出，“数据必须是原始的”。其次，二者的目的不同。政府信息公开主要在于保障公众的知情权。知情并不等于获得，获得也不等于可利用。因此，政府数据的共享开放是在知情的基础上，让人们获得和利用数据。

开放政府不完全等同于开放数据。2009年1月奥巴马宣布就职时，签署了《开放透明政府备忘录》，提出建立开放的政府方案，强调在数据时代政府的开放应当与技术结合起来，在开放政府的透明基础上，促进政府创新、合作、参与，提升其效率和灵活性等，并要求大多数联邦机构对其拥有的数据资源必须开放。同年，美国推出了世界上首个国家级的开放政府数据平台www.data.gov。2011年，世界各国建立起“开放政府联盟”，目前已有70多个国家成为会员，该联盟的其中一项工作就是开放数据。目前这些开放政府开放的数据主要是指官员财产、财政类数据等，却不包括其他数据，如科学数据、商业数据。

政府的数据开放，对内可以实现部门与部门之间的共享，有利于提高政府服务的办事效率，节约运营和管理成本。相对于按照传统专业划分的政府管理模式而言，政府部门之间达成一定协议和接口的数据资源共享开放，可以打破部门之间的信息壁垒，取代治理领域分裂的模式，形成完整的公共服务链条。各部门可以迅速获取所需的原始数据，而非处理过的数据，从而实现原始数据的关联分析，提高政府办事效率。这

样可以节约数据成本，避免浪费资源，最大限度地挖掘和利用数据的价值，实现各部门间的利益最大化。

另外，政府向企业、个人实行数据资源共享开放，能够提高政府的公信力，激发市场活力和社会创造力。应当在确保个人隐私、公众利益和国家安全的前提下，对外开放共享数据资源。促进社会各方对政府数据资源进行深度开发和增值利用，助推产业升级和经济转型，带动企业在技术、应用、商业模式以及跨界交叉方面的创新涌现，带动经济增长。要更好地满足公众需求，促进社会创新，提升公众和企业在政府数据资源共享和开放中的参与度。

为实现数据开放共享价值，推动政府数据一站式开放，要在建立政府数据开放平台和标准体系的基础上，依照相关法律法规，有序开放共享数据。通过推动云计算技术的发展，加快市场数据监管体系的完善，使政府和公共部门的数据资源的汇聚、整理、分析达到集中开放的标准规范，实现面向社会的一站式开放服务。

相对于数据一站式开放而言，公共数据资源的契约式开放也很重要，要以政府掌握的能够开放的数据为基础，成立公共数据开放平台，向国内所有的法人企业、创业者进行点对点的开放。对开放的规模、层次、展现形式、使用时限等以合同的方式进行事先约定，并根据相关法律和合同约定，对开放对象数据使用情况进行监管。通过契约式开放，吸引社会企业、社会法人参与，形成一种良性互动的循环，进而促进数据竞相开放，形成数据融合开放的格局。

（三）数据开放、流通和应用中的数据风险

数据开放中的数据风险是国家战略层面面临的主要威胁。数据开放让大数据时代的国家主权越来越相对化：一方面，数据的开放性和自由化，大大降低了政府对数据行为的管控能力，影响了各国之间的沟通、交流与合作；另一方面，数据开放使数据主权的争夺成为国家战略的制高点，给国家安全带来了严重的威胁。世界大国都在不断加强网络空间的战略谋划，抢占数据和网络空间的主导权，不断形成全方位的网络空间战略威慑力。

此外，数据开放过程中，数据的使用包括与第三方的合作，这使安全和隐私遭到极大挑战。开放的个人数据是具有身份标识的数据，加大

了个人隐私泄露的风险。匿名化技术带来的去匿名化技术让如何进行原始身份数据的保护成为一大难题。黑客攻击数据集聚的运营商不可避免地加大了用户隐私泄露的风险。数据的开放与保护是天平的两端，只注重数据的开放，不注重其安全保护，将会给世界带来一场灾难。因此，必须在开放中保护，在保护中开放。

数据流通中的数据风险包括数据采集、传输、存储过程中的安全问题。在数据采集过程中，可能存在数据损坏、数据丢失、数据泄露、数据窃取等安全威胁。[注]现代社会，网络服务商通过网站或客户端软件对网民进行个人数据采集的行为日益突显。然而，现行数据采集的范围和内容没有具体规定和统一标准约束，数据安全在采集过程中无法得到有效保障。此外，大部分互联网应用在利用网络进行数据采集时覆盖面过于广泛，无论其采集行为有无权限、用户是否知情，其采集行为都在进行。如果采集个人数据的应用服务商整体缺乏安全意识，没有必要的安全技术措施、人员和经费的投入，将无法有效保障采集信息的安全。

数据传输面临的安全问题包括机密性、完整性、真实性等。数据在传输过程中存在被窃取和篡改的问题，特别是在无线网络的传输环境下，网络传输中的数据安全问题尤为突出。在以支付宝、微信支付、苹果支付（Apple Pay）为代表的手机终端支付的案例中，用户的消费数据以及支付密码在流动传输过程中容易遭受黑客攻击，运营商如果没有足够的安全保障，容易造成数据泄露、账号被盗等安全威胁问题。斯诺登案件充分表明在网络传输过程中遭受数据窃听等威胁的现实性，窃听范围越广泛，人们的切身利益越无法保障。

数据存储的管理、安全问题突出表现为数据的管理权限不确定、访问控制问题以及存储能力不足等风险。在存储数据的管理过程中，应当在存储设备的所有权和使用权分离的情况下，确保对所存储数据的具体位置和权限进行管理。如何确保数据的所有权和访问权得到合法保护，隐私保护、数据加密、备份与恢复等，都成为数据存储过程中急需解决的重点和难点。

针对网站等网络应用的攻击和漏洞呈现出批量化、规模化的显著特点，数据应用的安全问题主要表现为账户攻击愈演愈烈，漏洞发现和利用的速度越来越快，第三方代码托管平台被攻击，企业信息、个人隐私得不到有效保护等。从公安机关统计的数据来看，近几年网络违法犯罪活动呈快速增长态势，一些传统类型的犯罪也更多地利用和针对互联网

实施，不法分子利用各种手段窃取、贩卖公民个人信息，从事各种违法犯罪活动，遭到窃取、贩卖或泄露的信息涉及金融、教育、医疗、保险等重要部门和行业。

第二节 块数据风险源

（一）安全风险意识

安全是永恒且没有终点的话题，是人类最重要、最基本的要求。每种新技术的诞生总是伴随着更多种类、更高层次的安全问题，块数据的发展亦是如此。所谓安全风险意识，是行为主体对各种有可能带来危险的情况的警觉和防范的心理状态。当行为主体在意识中对某种事件和状态有警觉时，就会主动调整自己的行为轨迹以规避风险。

数据安全问题辐射面极广，与人们的生活息息相关。新兴社交平台的出现、电子商务的发展，在很大程度上改变着人类的生产生活方式，但在使生活变得更便捷和人性化的同时，也带来了很多的安全隐忧。数据技术的快速发展应用与安全意识之间的脱节，是数据安全问题产生的主观原因。目前人们的数据安全意识较为薄弱，绝大部分人都认为自己的网络环境较为安全，并不认为存在很多的网络攻击和安全威胁。

提升人们的数据安全意识，形成数据安全观是规避数据安全问题的有效手段。强化安全意识的前提是辩证地看待数据与人的关系。应当在数据空间中建立一种新的安全观，调整安全风险认知。从人类的头脑开始改变，通过重塑我们思考、评估、判断和决策的方式，恢复人类本身的自适应能力以应对数据的变迁。此外，在建立参与机制的各方中，人人都应具备相应的安全意识和风险意识，共同推动数据安全的可持续发展。

（二）关键信息基础设施的安全可靠性

一个国家和地区的关键信息基础设施主要包括低功耗新型硬件技术、基于大数据应用的下一代通信终端、一体化数据存储处理服务器以及软硬件设备一体化的大数据产品和大数据硬件设备等。由于网络信息技术的快速融合发展，国家关键信息基础设施的边界已经扩展到一些需要进行网络安全保障、使用信息技术支撑重点行业或领域重要业务正常

运行、事关国家安全的设施。

良好的关键信息基础设施建设可以降低创业成本，为大众创业、万众创新提供良好的环境，增加公共产品和公共服务，对信息消费、拉动投资以及促进新型工业化和信息化具有决定性意义，是经济社会发展的加速器。在国家信息安全战略层面，某个基础设施或组件的关键性意义在于，它所担任的角色或发挥的功能的重要性。关键基础设施存储或传输的信息数据大量集中，甚至极其敏感，如金融、地理、人口、商业数据交易平台的关键数据资源，一旦这些设施遭到破坏，会对国家安全、社会稳定和公众安全产生严重影响。基础设施安全防护能力不足会引发数据资产失控。基础通信网络关键产品缺乏自主可控能力，成为块数据的安全缺口。

（三）黑客与漏洞

黑客泛指擅长信息技术的人群，分为正派和反派两大类。其中，正派的黑客利用自己所掌握的知识和技能在网络系统中找出漏洞，目的是帮助管理人员修复漏洞并不断完善系统，人们也称他们为“红客”，他们坚守着《黑客守则》；而反派的黑客则运用各种黑客技能入侵系统，窃取机密信息，甚至对系统发动毁灭性的攻击。随着数据技术的发展和应用，黑客逐渐成为非常庞大的群体。深入黑客行为的本质，从目的出发可以将黑客行为分为窥探、恶作剧、炫耀“技能”。从行为本身来进行分类的话，黑客行为可分为学习技术、伪装自己、发现漏洞、利用漏洞。

黑客效应带来的安全威胁是无法想象的，甚至是致命的。移动电话系统和以计算机为基础的通信设施的结合，更有助于黑客复制微芯片，一旦其中某个系统遭受入侵，整个城市的电信系统就可能瘫痪。一些重要部门的网络系统管理人员的个人电脑上如果有所在部门的重要机密信息，黑客就能顺藤摸瓜窃取到更多机密。黑客还能通过无线电、移动电话等对重要部门进行网络渗透。这些都将严重地威胁个人信息安全、社会稳定，甚至是国家安全。

黑客通过技术薄弱环节及漏洞入侵。如果核心操作系统、关键技术设备、大型数据库等关键信息基础设施都采用国外的产品，则极易带入许多嵌入式病毒，比如可恢复密钥的密码等，这样会遗留众多安全漏洞以便黑客攻击，同时技术公司还能够利用这些设备的“后门”来窃取国家高风险数据，甚至还可能发动网络攻击破坏国家的关键基础设施，直接

威胁国家安全和社会稳定。漏洞是黑客的重要线索，他们的目标就是利用技术漏洞进行攻击或是对漏洞进行修补，并从大量的漏洞中筛选高价值、可利用的漏洞进行试验。因此，试验的前提是发现漏洞，通过漏洞黑客可以获得系统信息、入侵系统、寻找与之关联的下一个目标。如果数据安全技术手段的发展和创新速度无法跟上大数据井喷式的增长，则数据的安全防范会更加困难，威胁将日益凸显。

（四）数据恐怖主义

数据恐怖主义源于网络恐怖主义与传统恐怖主义。网络恐怖主义最早由美国加州情报与安全研究所在1997年提出，即网络与恐怖主义相结合。网络恐怖主义与其上位概念——恐怖主义之间既有联系又有本质上的区别。相较于传统的恐怖主义，网络恐怖主义并不会直接采取暴力行为造成人员伤亡，而是通过更为隐蔽的网络信息技术入侵受害方系统，进行操控和攻击。因此，定性上通常认为两者的本质是智力和暴力的区别。

大数据时代，网络已经渗透到国家发展的金融、军事、科技、教育等方方面面，成为一个国家正常运行和赖以生存的基础。恐怖分子同样充分地认识到了这一点，并利用网络信息技术使得恐怖活动数据化，更具隐蔽性和破坏力。大数据为恐怖主义提供信息、技术和空间支持。由于数据已经成为国家的基础性战略资源，就更容易成为恐怖分子的目标，尤其是技术人员不断突破的数据技术，反而成为恐怖分子发动攻击和破坏的武器。大数据的开放提供了海量的数据资源，让数据恐怖活动变得更加精准。

数据恐怖主义是网络恐怖主义在大数据时代下发展的更高阶形态，它具有网络恐怖主义的典型特征，但是更加具有渗透性和摧毁力。随着数据的重要地位日益凸显，政府各相关部门、企业均越来越重视对数据资产的管控，高危敏感数据很可能与国家安全、经济稳定息息相关。因此，未来的网络恐怖主义最为可能的攻击目标就是关键敏感数据，用以操控政府机构、关键基础设施和公共安全。

对数据的依赖日益加深，威胁也就越发严峻。如果不加以有效防范，数据恐怖主义将进一步冲击国家主体地位、威胁国家意识形态，从而威胁国家政治安全；甚至还会通过修改重要的财经数据来扰乱银行系统，诱导做出错误的经济决策，导致国家经济体系的崩塌；抑或入侵国

家防御系统，通过使国防关键性杀伤武器自动销毁来削弱军事力量等。因此，为了维护世界和平、稳定发展，凝聚各方力量共同对抗数据恐怖主义变得至关重要。

（五）法律缺失与滞后

数据是未来社会发展的基础性资源。数据领域的法律政策所面临的挑战，正随着数据的发展日益凸显。如何平衡数据利用与数据安全，涉及数据利用与数据资源本身安全之间、数据利用与隐私保护之间，以及数据利用与国家、社会和个人之间的安全关系，已经成为一个世界性难题。

大数据时代的快速变迁加剧了数据立法的滞后性。一方面，大数据发展的不确定性，使对于立法关系中“主体的行为”的解释、预测和控制变得异常困难，不能有效地指导立法实践；另一方面，数据立法的理论研究远远落后于数据技术的发展及实践的变化，存在很多研究盲区，立法与社会需求脱节。没有系统性地构建覆盖整个大数据网络领域的法律体系，导致研究很难深入。

此外，当前急需在数据确权的问题、数据的价值衡量与定价问题、数据作为重要资源自身的安全问题上明确法律边界，要对在贸易过程中数据的跨境流动和限制问题，以及数据向特定平台汇聚形成的数据垄断提出明确的法律界限，划定大数据时代的新责任主体的权利义务归属等，尽快解决大数据领域的法律缺失与滞后问题。

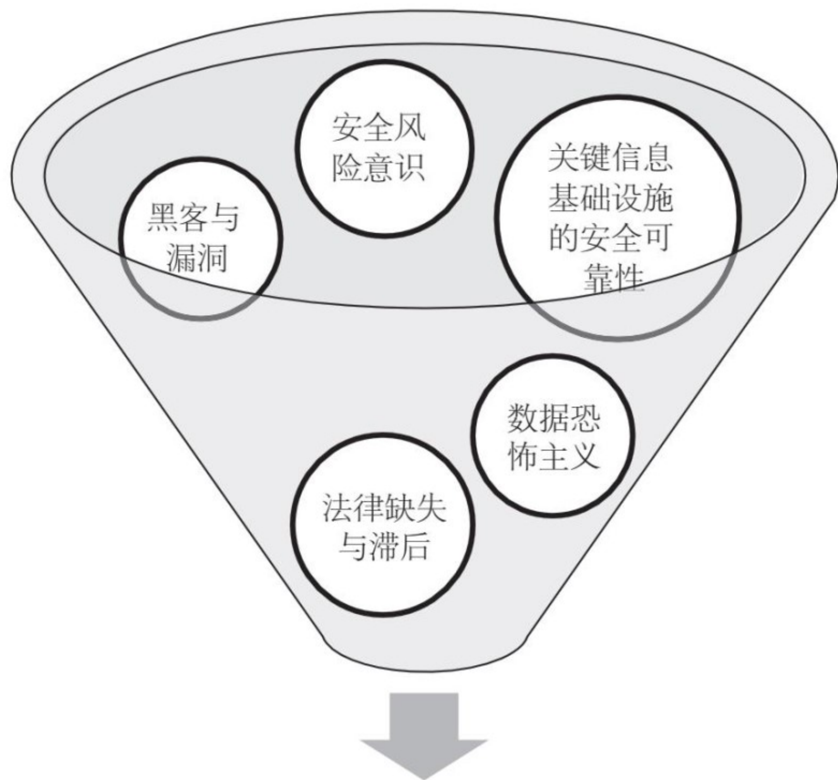


图10–1 块数据风险源

第三节 块数据安全防御

（一）重点领域的信息保护

加强对关键行业领域重要信息系统的安全保护，保障网络数据安全。对于国家重点系统、重要部门的网络信息系统和基础设施应当严格执行等级保护的有关规定，加强监督检查，实行更为严格的信息安全等级保护管理制度。军队、公安、金融等领域的数据要从物理上、技术上和外界完全隔离，杜绝从外部入侵的危险。

对于数据行业，要制定严于一般行业的管理制度，加强对于从业人员的教育、管理以及内部防御。采取更高规格的信息安全等级管理保护措施，明确不同人员相应的管理权限，加强监督制约，实现层层管控，时刻严密管控网络信息安全动态，及时有效处置网络信息安全问题，保

障网络信息安全。在国家层面加强对数据的掌控，规避数据聚集利用给国家安全带来的风险，在法律层面对信息资源制定保护措施。

（二）涉及国家安全稳定领域的产品和服务

互联网是数据存在的最重要载体，规划设计具有中国特色的自主可控的下一代互联网是维护国家数据安全的重要内容。为实现下一代互联网的自主可控，关键在于加强互联网基础设施和关键硬件设备的所有权和控制权，自主掌握基础软件技术的开发和应用，掌控互联网传播领域中代表国家意志和主流意识形态的话语权，保障国家、组织及个人在全球互联网通信中的自由和信息安全。

在推进下一代互联网等数据系统建设时，要实现建设、应用、管理与安全的统筹规划，同步做好顶层设计。数据安全问题与块数据发展的全过程是相伴相生的，从数据系统的建设伊始到其应用、管理，再到维护、改进、完善的全过程都必须要充分考虑数据安全问题，同步构建有利于数据全过程保护的安全性架构和安全性平台。

维护国家数据安全的关键是掌握互联网核心技术，互联网核心技术受制于人是我国互联网领域主导权较弱的关键原因。要把互联网核心软硬件技术列为国家战略性重大工程，在网络软硬件领域，开展基础性、原创性研究，力争取得关键性突破，掌握互联网这一数据发展最核心领域的主导权。同时，要争取在数据安全基础理论、云安全、交易安全、人工智能等数据安全重点领域掌握更多的核心技术及话语权。

（三）关键信息基础设施的安全可靠水平

关键信息基础设施一旦遭到攻击和破坏，就会对国家安全、经济稳定和公众安全产生严重的影响。因此，在宏观层面，可以从业务连续能力、设备的自主可控、敏感数据安全维护以及关键信息基础设施的主体责任4个方面来保障关键信息基础设施的安全可靠性。

业务连续能力是关键信息基础设施安全保护中需要解决的首要问题。1998年美国签署的《关于保护美国关键基础设施的第63号总统令》中要求：“采取所有必要的措施来迅速减弱关键基础设施，尤其是信息系统在面临物理和信息攻击时的任何重大脆弱性。”2003年2月美国发布的《网络空间安全国家战略》中所列的网络空间安全的战略目标包括：“降低国家在网络攻击前的脆弱性”和“缩短网络攻击发生后的破坏和恢复时

间”。《中华人民共和国网络安全法（草案）》从规划建设、使用管理、人员配置、容灾备份、应急处置等多个方面做出规定，来规范关键信息基础设施的业务连续能力。

解决块数据安全问题的根本途径是实现关键信息基础设施的自主设计制造。“棱镜计划”的实施方式就是美国国家安全局通过各大技术公司的产品“后门”进入各国的信息系统，并获取相关信息。这就意味着如果我国的关键信息基础设施还继续大量使用未经安全认证的国外设备，将给我国的业务系统和数据安全埋下风险隐患，对国家信息安全构成潜在的威胁。因此，必须基于国产化和自主知识产权建立安全可信的信息系统，其中最核心的就是关键信息基础设施中关键设备的自主可控，实现国家对主要信息产品、设备和技术的可控管理与使用。

维护敏感数据安全应进一步明确哪些数据可以在国外的服务器上储存，哪些数据必须在国内的自主服务器上存储，限制重要数据出境。俄罗斯最新出台的一项法规规定，所有采集俄罗斯公民信息的互联网公司都应当将这些数据存储在俄罗斯国内，该法律的生效时间为2016年9月1日。我国关键信息基础设施的运营者应当将运营中采集和产生的重要数据存储在我国境内，以加强对于国家和公民的个人数据保护。

关键信息基础设施责任承担主体分为两个部分：国家和网络运营者。在关键信息安全基础设施安全防护当中，国家的主要职责和义务是监督和指导。在实际运营过程中，某些关键信息基础设施的安全问题是网络运营者依靠自己的能力无法解决的，这种情况下必须借助外部力量进行保障。国家相关部门要为网络运营者提供技术支持和援助以应对网络安全事件的应急处置和恢复。此外，针对重要的关键信息基础设施建立一套应急备份和灾难恢复机制，是国家层面的责任和义务。

网络是信息的主要获取和传播渠道，运营者要负起责任，保障网络自身的安全可靠。同时，确保使用者个人信息的安全，防止泄露；建立网络信息安全投诉平台，及时处理有关网络信息安全的举报等。此外，还应明确使用者在关键信息基础设施安全保护中的义务和责任，以保障关键信息基础设施的安全。

（四）数据安全标准体系和评估体系

随着数据安全问题的日益严峻，各国和国际标准组织纷纷出台数据

安全相关标准指南，比如美国国家标准与技术研究院（NIST）发布了用户身份识别指南；国际电工委员会（IEC）和国际标准化组织（ISO）制定了公共云计算服务的数据保护控制措施实用规则，而我国在数据安全标准建设方面的探索才刚刚开始。由于大数据的应用范围比较广泛，大数据的安全标准体系也比较复杂，应该重点研究制定基础标准、技术标准、应用标准和管理标准等共性安全标准。针对个人隐私、电子商务、国家安全等重点领域和安全问题多发领域，率先研究适用的数据安全标准，切实保护国家数据安全以及公民、法人和其他组织的数据安全。数据安全问题贯穿数据处理始终，要研究形成覆盖数据采集、存储、挖掘、公开、共享、使用、管理等数据处理全过程的安全标准体系。

数据安全评估体系的建设也是确保数据安全的重要保障，近年来美国、欧盟等开展了数据安全评估工作，如美国TRUSTe公司为3 500多家门户网站和公司提供隐私认证服务，欧盟数据跨境流动安全评估成为评判数据能否转移的重要依据。针对数据平台及服务商等重点对象，做好可靠性及安全性评测、应用安全评测、监测预警和风险评估。完善网络安全评估监测体系，实时监测和感知网络安全威胁，防御网络攻击，提升对大规模网络攻击威胁的发现和应对能力。借鉴欧盟的做法，加快开展数据跨境流动安全评估，强化数据转移安全的检测与评估，确保数据在全球流动中的安全。

（五）数据安全监测预警系统

网络攻击、数据泄漏和数据窃取是造成数据安全问题的三种最主要方式。实现数据安全预警的关键就是要把握重点，堵住造成数据安全问题的漏洞，防患于未然，将数据安全问题消灭于萌芽状态。

防攻击。随着网络攻击工具与攻击技术的新发展，使用互联网进行业务运营的机构面临新的风险。网络攻击是一种利用互联网存在的安全缺陷与漏洞，对网络系统的软件、硬件、系统中的数据进行一系列攻击的总称。因此，防攻击就是指当恶意攻击者对应用系统及数据库进行攻击或恶意操作时，管理人员及系统能够利用有效的手段阻止恶意行为的发生，减少信息泄密的发生次数，避免对企业和信息安全造成影响。

防泄漏。当前导致数据泄露的一般分为业务和技术两个层面的原因。注我国数据泄漏的根源在于引进为主、自主为辅，缺乏话语权。在我国最核心的部门，如军队、军工、政府、金融行业、保险行业、电

信行业等之中，就有高达80%以上都在使用国外的数据库产品，这些部门也是我国数据泄露最集中、最严重和泄露威胁最大的重灾区。[注]数据安全预警体系中的防泄露是指通过应用系统及数据库的风险评估，发现可能存在的信息泄密点，并进行安全加固。随着应用系统及数据库的不断升级，机关部门应及时了解并掌握应用系统及数据库自身是否存在着安全隐患，尽可能避免漏洞对外发布，减少信息泄露带来的危害。

防窃取。网络窃取是指数据信息在未经加密的情况下便以明文形式传送，而入侵者在信息传送环节如网关和路由器上截获信息的行为。入侵者通过对数据包进行多次盗取和分析，可以发现信息的规律和格式，进而获取信息内容，使网上传输的信息产生泄密风险。数据安全预警体系的防窃取是指对于任何人在任何地点任何时间进行的企业信息访问行为能够具有追溯手段，为发生的信息泄密事件提供查询工具，方便维护人员及管理员进行事件追踪和定位，并为事件的还原提供有力依据。

通常，攻击、泄露、窃取三者相互交叉、相互融合、同时发生，是一系列入侵行为，它们危害极大，后果难以估量。可以从源头、环节、系统三个管理体系上进行综合考虑并形成溯源机制，也可以从网络安全、应用安全、操作系统安全三个方面进行安全技术保障，更要加强立法与处置研究，从严、从重制裁此类犯罪。

（六）数据安全保密防护体系

构建数据安全保密防护体系，应当将管理与技术手段相结合，从管理层面和技术层面用规范的制度进行约束，全面提升数据安全保密及防护的综合能力。

从管理层面看，数据安全保密防护体系大体可以分为制度管理、资产管理、技术管理与风险管理等方面。制度管理是指制定、审查、监督与落实数据安全保密制度；资产管理是指对涉密人员、涉密场所、重要数据资产的备份恢复以及网络与计算机的管理等；技术管理是指对存有泄密隐患的技术进行检查，对安全产品与安全系统中涉及的技术进行测评，对各种泄密事件进行技术取证；风险管理是指评估与控制安全保密风险。

从技术层面看，数据安全保密防护可以依靠的技术手段主要包括电

磁防护技术、通信安全技术、信息终端防护技术、网络安全技术等。通过技术手段从不同层面保护数据系统与数据网络，保障数据及数据系统的安全，提升数据系统与数据网络的安全可靠性与抗攻击能力。

（七）数据安全能力

保障和提升块数据安全的能力主要包括态势感知、事件识别、安全防护、风险控制以及应急处置5种能力。

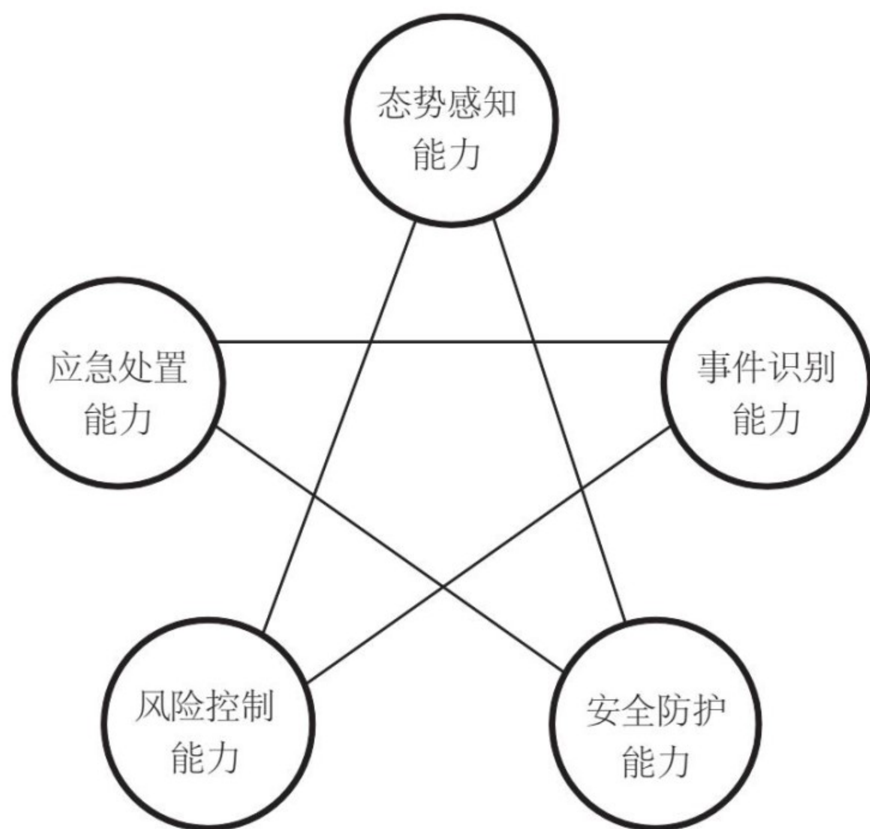


图10–2 保障数据安全的5种能力

态势感知能力。提高网络和大数据态势感知能力，即通过对安全要素进行综合分析，从而评估网络与信息安全的总体状况，并预测其变化趋势。提高态势感知能力不是孤立地研究网络信息安全事件，而是要综合多方警报和流量信息，通过聚合、关联、融合、归并等方法建立定性或定量描述的指标体系，达到精确感知态势的目的。块数据技术特有的海量存储、并行计算、高效查询等特点，为大规模网络安全态势感知

技术的突破创造了机遇。

借助大数据分析，对成千上万的网络日志等数据进行自动分析处理与深度挖掘，对网络的安全状态进行全面分析和综合评价，感知网络中的异常事件与整体安全态势。

事件识别能力。块数据的核心在于精准“预测”。在攻击事件发生前，通过对网络攻击的海量数据进行采集、分析与计算，发现网络攻击的异常行为和规律，有效识别攻击源和网络的风险点，准确预测网络和数据安全事件的发展趋势，使网络攻击无所遁形。在攻击事件发生后，通过数据溯源机制标记和定位大数据应用的各个环节，这样便可以及时准确地定位出现问题的环节和相关责任者。

安全防护能力。构建数据各相关主体的安全防护能力，需要做好准备工作与保护工作，提升防范水平，以应对入侵者的攻击，使数据主体避免危险、不受侵害、不出现事故。此外，应当全方位地保护数据应用和处理的各个环节，例如采集、存储、挖掘、开放与应用等环节。保障数据安全，不仅需要安全技术，还需要专业的安全管理，因此建立起针对大数据特点的安全管理、安全评估模型也是安全防护能力建设的重要内容之一。

风险控制能力。即通过风险识别、确定和度量来制订、选择和实施处理方案，从而减小甚至消灭风险发生的可能性，同时降低损失。风险控制包括事前、事中和事后三个阶段，风险事件发生前的首要风控目标是使潜在损失最小，而事中和事后的重点则是使实际损失降到最低。网络和块数据系统中的风险往往具有隐蔽性、影响大、波及广等特点，因此要把风险控制的重点放在风险发生前的识别和控制上，要研究开发数据采集、传输、存储、利用、开放等全过程的风险识别技术，真实、客观、公正、全面地反映风险程度，制订风险管控措施和解决办法，将可能出现的风险控制在可以预防的范围内。

应急处置能力。应急处置是数据安全能力建设的最后一道防线，构建和完善包含应急响应、联动处置、数据恢复以及数据灾备等子系统的应急预案体系是重中之重。应急响应组织是根据事先对各种可能情况的准备，在数据安全事件发生后的响应、处理、恢复、追踪的方法及过程。在数据安全应急处置过程中，部门之间要形成有效联动和协调配

合，统一指挥、分级负责、专业处置。数据恢复的重点是通过技术手段，对保存在存储设备上丢失的数据进行抢救和重新找回。数据灾备包括数据备份和数据容灾，是数据安全和可用性的最后一道防线，其目的是为了在系统崩溃时能够迅速恢复数据，但仅有备份不够，真正的数据容灾应该能够弥补传统备份的不足，在灾难发生时可以及时恢复整个系统。

（八）隐私与个人信息管理和保护

隐私与个人信息管理和保护应该贯穿在数据采集、处理、交易、应用等全过程的每个环节，其关键是要规范各利益相关主体的行为。

在数据采集阶段，主要涉及个人、政府、企业三方主体。对于个人而言，最重要的是培养隐私和个人信息保护意识。一方面，个人要善于运用法律武器保护自身的权益，同各种违法犯罪行为做斗争；另一方面，要增强个人的法制意识，普及隐私信息安全的法律知识，使个人懂得上网时要遵守法律和社会公德，不能以自己的自由侵犯他人的权利。对于政府与企业而言，要规范公共部门和企业数据采集方式，完善个人隐私和信息保护的相关法律法规，完善公共部门和企业共同遵守的隐私保护条款，明确政府相关部门企业、行业、网民在数据社会要承担的法律和社会责任，各负其责。

在数据处理阶段，主要涉及政府、企业、行业组织三方主体。要加快建立个人数据处理审查机制，充分发挥政府和行业协会的作用。政府要切实担负起对企业等数据运营主体的审查责任，对违规部门和企业要严格处理。从国家层面制定匿名（化名）处理标准规范，使之成为各数据处理部门、企业共同遵守的规则，提高个人隐私和信息的保密等级。建立行业个人隐私信息审查协会，专门从事个人隐私和信息保密情况的审查工作，充分发挥第三方客观、公正的作用，为裁决部门提供参考。

在数据交易阶段，由于涉及多方参与，极易出现个人隐私和信息的泄露。为此，一是要建立个人数据销售许可机制，通过颁发可转让许可证、采用拍卖授予方式、明确许可销售数据类型、建立许可退出机制，使数据交易公开、透明，避免出现黑市交易。二是要建立个人数据流转登记机制，使数据的流转更加可控，交易过程更加有序。三是建立个人数据跨境流动审查机制，在跨境数据交易中，要进行严格审查，避免隐私和个人信息在更大范围内泄漏，造成更大损失。

在数据应用阶段，一方面，要建立多元参与的个人数据隐私泄露举报机制，加强网络社会综合治理，动员企业、个人、媒体等多方参加数据社会的监督管理，共同参与数据安全治理；另一方面，建立个人数据隐私泄露溯源机制，一旦发生隐私和个人信息泄漏，要能找到责任主体，使泄漏者承担相应法律、社会责任，同时发现薄弱环节，弥补制度缺陷。

（九）审慎监管和保护创新

监管与创新是对立统一的矛盾体。一方面，监管刺激创新产生。对于数据发展的监管既是创新的制约，也是创新的诱发因素。监管为数据企业的创新发展提供了公平公正的环境，推动市场健康发展。但是过度监管会制约创新的发展。由于过度监管增加了数据企业机构的运营成本，降低了其赢利能力，当监管的约束大到回避它们便可以增加经营利润时，经营机构便有了创新的动力。另一方面，创新促使监管不断变革。创新的出现在一定程度上对数据监管体系提出了新的挑战，对监管模式和方式的改进提升都具有推动作用。

因此，既要审慎监管，又要保护创新，平衡二者关系。正确把握审慎监管与创新之间的平衡点，让两者相互协调、健康发展，实现从监管与创新到再监管与再创新的良性循环发展。要做好以下几项工作：一是不断提升数据发展的创新层次；二是加强数据发展创新过程的监管；三是完善数据发展监管协调机制；四是加强数据发展监管的国际合作与区域合作。

（十）数据立法与数据伦理

鉴于数据立法的缺失与滞后，世界各国都在加快探索数据立法，我国在数据立法建设方面也有了明确要求。2015年，国务院颁布的《促进大数据发展行动纲要》要求加快法规制度建设，明确指出要研究制定数据开放保护等方面法律制度，制定信息资源管理办法，推动网上个人信息保护立法工作，保障网络数据安全，研究数据资源权益相关立法等。在鼓励地方立法“先行先试”方面，《贵州省大数据发展应用促进条例》作为全国第一部地方性大数据法规，为其他地区提供了先行先试的经验。

对大数据交易与数据权归属问题的确认目前仍然是法律空白。数据

作为一种资源，是产权的一种表现形式，具有确定的、清晰的所有权，需要由法律进行保障。而作为立法主体的活动准绳，立法原则的明确是重中之重。从目前来看，数据立法应当重点明确保护公民权益、保护国家权利、国家安全和公共利益、促进大数据产业发展等基本原则，为将来的立法实践提供指导。

既然数据权成为一种法定权利，研究数据资源是否具有财产性就成为解决数据确权归属，并对其进行保障的重要前提。从市场实践的角度出发，个人数据商品化充分表明其财产性质；从法律角度看，个人数据权利的来源是人格权。目前，关于人格权与财产利益之间的关系仍然存在很大争议。

不论是物权还是数权，都有对应的所有权人。以个人数据为交易对象，个人数据的所有权人即数据主体当事人。而在个人数据的基础上，对数据做出匿名化处理后形成了数据集，企业又成了数据的控制者，对匿名化数据集享有有限制的所有权。个人与企业可能是重要的数据权利主体，但不囊括所有的权利主体。因此，关于如何划分数据所有权人的问题，还需要进一步厘清。正如奇虎360公司董事长周鸿祎在2015贵阳数博会上所讲到的，所有大数据都是由用户产生，它被传到各个互联网公司的云端服务器或者传到政府的云端，其所有权属于政府还是属于互联网公司，就是一个重要的法律问题。就此，他提出了大数据时代的信息安全三原则，即信息是用户的个人资产，其所有权属于用户；互联网公司和政府通过服务换取用户的数据，其对用户数据的使用必须经过用户的授权和认可；作为存储用户数据的主体应该对用户数据提供最安全的保障。

企业对于匿名化数据集是否应该享有有限制的所有权，是数据立法的重点和难点。或许可以从权利的优先顺序考虑；从个人信息保护法的原则进行考量；从匿名化技术及大数据发展趋势综合解剖，从而为我们找寻答案提供一些思路和研究基础。同时，还要从立法层面界定数据清洗与数据匿名的不同。数据清洗的结果并不必然实现数据的匿名化，在涉及所有权问题时，应当强调只有经过充分匿名化的数据，企业才对其享有所有权。

马克思曾说过，技术一旦被人发明，就如同存在一个某种自主性的“他者”对人自身产生影响，发挥着种种反主体效应。数据安全不只是技术问题，块数据催生的也不只是技术发展。大数据时代人类社会所面

临的最大挑战是数据伦理和数据道德问题，是数据世界的不确定性、不可预知性、不可控性，是人性在数据世界面前的无限放大，由此造成的巨大隐忧。这种隐忧不仅包括现实的个人隐私泄露、网络犯罪等问题，还涉及更长远的数据伦理、数据人格和数据人性问题，不仅包括人与人的数据伦理问题，还包括人与机器人的数据伦理问题，这将是大数据时代直接影响人类发展的重大关键问题。就像克隆技术的发展已经挑战了人类的伦理禁区和道德底线一样，“数据＋科技”的结合或许会给人类带来更大的伦理道德危机。例如，新西兰奥克兰生物工程研究所已经研制出了一个名为“艾可丝宝宝”（BabyX）的人工智能人，她的“大脑”是模仿真正幼儿神经系统的计算机程序，具备学习和认知能力，研究者的终极目标是创造一个能思考、像人类小孩一样会学习的机器，最后长大成为人工智能成年人。不久的将来，真正具有人类思维和情感的高智能机器人真的有可能出现，届时，数据世界中的虚拟人类与现实社会中的人类之间的伦理关系如何处理，将成为人类必须面对的全新伦理问题。

基于此，在强调立法重要性的同时，丝毫不能忽视社会伦理道德体系建设在数据安全领域的重要意义。法律法规和伦理道德是社会的经纬。数据社会中，法律是伦理道德的底线。加强数据立法的同时，更需要建立一套普适的数据伦理与数据道德体系。

数据伦理与数据道德本质上要以人为主体，它内在体现的必然是一种以人为本、尊重人的尊严和价值、维护人的地位和权利、实现人的发展和追求的人本意识。数据本身是中性的，数据运用是否得当、是否合理，这是技术无法解决的。数据安全问题，在很大程度上应该用伦理道德来规范和约束。我们需要做的是理顺人类世界与数据世界的伦理关系，共同迈向大数据时代的人类命运共同体。

---

1. 傅杰：《消除数据孤岛，摆脱企业大数据困境》，2015年1月13日，http://server.51cto.com/BigData–463827.htm。
2. 互联网新垄断，《财经国家周刊》，2011年9月19日，http://news.sina.com.cn/c/sd/2011–09–21/155523193713.shtml。
3. 互联网新垄断，《财经国家周刊》，2011年9月19日，http://news.sina.com.cn/c/sd/2011–09–21/155523193713.shtml。
4. 《“无边界”组织与“信息流”》，凤凰网，2010年，http://www.rxyj.org/articles/264331.html。
5. 陈谭：《大数据时代的国家治理》，北京：中国社会科学出版社，2015年。
6. 迪莉娅：《大数据环境下政府数据开放研究》，北京：知识产权出版社，2014年。

结语
块数据场景应用八大预测

在未来，也许就是不远的50年、20年乃至10年以后，以块数据理论为指引，伴随着物质的极大丰富和技术的极大发展，社会的生产、生活乃至思维模式将发生极大变化，全球将会迎来一场持久而深远的数据化革命。在已知世界的尺度内，万物都将可能被量化、记录和分析，事件、科技、生产、生活的发展趋势将变得可以被研究和预判，世界将会以数据的方式重新表达、关联、解读和进化。也许，某一天，在一个可以预判发展趋势的社会里，我们将重新回到“计划”时代，精准、定制、个性、公正将成为时代的核心特征，块数据将成为社会形态的表现形式，创新创业将成为人们生活的基本需要。完全可以预想，以下诸种形态说不定某一天就会悄然出现……

（一）DNA穿戴城市。 在衣裤、眼镜、手表或者发带等贴身物件上装上微型穿戴式DNA识别仪，这是一个看起来很普通，但可触摸、能连接人体的身份识别装置，能够在以毫秒为单位的时间内响应、读取、传输数据，与DNA数据库进行个人数据实时比对验证，并将DNA数据和个人财务账户、健康管理账户等绑定，为个人工作生活提供极大便利和安全保证。与传统的虹膜、唇纹、指纹等识别装置相比，该设备具有自主性，客户穿戴与否全凭自己主观意愿，不会侵犯个人隐私，且无须像芯片那样植入人体，造成物理伤害和内心伤害。如果你看过《黑客帝国》这部电影，想必对这一设定会非常熟悉。人体内植入芯片，首先意味着人被异化，这恐怕是每一个个体所难以容忍的。同时，这一设备还具有普适性。每个人都具有独一无二的DNA，且全世界范围内DNA的鉴定标准是基本统一的，因而识别仪能够在各种场合、各个国家通用。这就为身份数据的识别和标准的统一创造了条件，实际上是让人真正成为块数据环境下原点数据的入口。这一设备的原理并不复杂，只是要求它具备点对点验证功能，能够确保识别每个人DNA的唯一性，进而确保与之链接的账户、门锁等的安全。

穿戴式DNA识别仪的以上特征使它能够在多个场景中使用，并具有高度安全性、便利性和精准性。完全可以想象，当你外出的时候，你就

不再需要带手机、钱包、钥匙、身份证、银行卡等证件，只要穿上附有DNA识别仪的衬衣或者袜子，戴上DNA识别手表，走到地下车库，车门将在你靠近的时候自动打开；平常上班不用在指纹打卡机上按下指纹签到，参加高端活动时也不需要走安检通道，DNA识别仪在你走到入口之前已经自动完成了身份识别；午间去餐厅就餐，吃完就可以离开，不用担心支付的问题，DNA识别仪绑定的银行账户会自动结算，你根本不需要在这些事情上耗费时间和精力。有理由相信，穿戴式DNA识别仪将在块数据时代成为实现人机交互的核心设备之一，会颠覆现有的身份证、各类虚拟支付手段、信用保证等。

（二）云出行。借助融合了城市的经济数据、地理数据、消费数据等各类条数据的智能化城市管理块数据平台，通过无人驾驶汽车、无人机等智能设备，完全可以实现市民的全免费、全智能、全个性化出行。与传统的出行方式相比，云出行将实现出行需求和线路的实时化调度。以城市管理块数据平台为中心，每一位市民、每一个交通工具及与城市日常管理相关的条数据全部汇集在平台上，通过先进的算法，对市民要去哪里、去干什么、怎么去进行预先设计和规划，对相应的工具、路线、时间等资源都将在整个城市的范围内进行按需分配。更让人愉快的是，云出行还将实现交通工具的智能化驾驶。现有的汽车、高铁、飞机将会被小型化、智能化、共享式的无人驾驶“交通舱”取代，通过尖端通信和深度学习等方式，自行选择路线、速度，自行避让障碍物，并与其他“交通舱”进行数据共享和交流，共同确保通行过程最大限度的安全、高效、精准和生态。

比如，当你有出行需求时直接通过DNA识别设备与城市管理块数据中心联系，过一会儿，离你最近的空闲“交通舱”将出现在你面前。特别是，如果你住在高层建筑中或者地处道路交通不便的地方，可飞行的“交通舱”将自动上门，实时提供交通服务。“交通舱”将根据你的目的地和时间安排甚至是心情自行计算最优的出行线路，你只需要进入“交通舱”就可以了，至于你在舱内是想坐着、躺着还是站着，完全按照你自己的意愿，你不仅能够从方向盘上解放出来，而且将享有高度虚拟化的工作和娱乐载体。你可以休憩，或利用全息系统阅读、听歌以及游戏。总之，在智慧出行场景中，红绿灯、驾驶员等都将消失，交通拥堵将不复存在。

（三）长寿城市。基于个性化的适宜的生态环境，安全且富有针

对性的食品供应，无诊疗式慢病管理，科学高效的突发性应急指挥调度系统，较低的工作、生活压力，构建人均寿命远超过同等收入水平城市的新型城市。与一般城市相比，长寿城市的构建理应基于以下要素：一是对生态环境的精细化、人性化管理和保护。每一条街道、每一个小区甚至每一栋楼的生态环境指标都将通过DNA识别仪绑定，根据每一个人周围的生态微环境指标提醒、预警其生态状况，并进行微调整，让每一个人始终处于最适宜自身状况的生态环境中。二是食品监管与配送的智能化。通过块数据手段，把每个个体的饮食喜好数据、DNA数据、身体健康数据、食品安全监管数据等进行融合分析，精准地识别差异化需求，开展定制化、小批次的生产、供应和配送，最大限度地把每个人的主观喜好和身体对营养成分的客观需要结合起来，并保证食品的安全无副作用。三是特殊DNA群体的无诊疗式慢病管理。随着医疗块数据的日趋完善和丰富，医生能够及时甄别、发现每个人身体健康方面的细微变化，通过精准提出饮食调养、健康休养建议方案等，提前进行微调整、微修复，真正实现通常所说的“治未病”，避免诸如心肌梗死、脑溢血等突发疾病导致的猝死。四是突发性事件的自流程化应急处理。借助城市突发事件应急管理块数据指挥中心，事前就能够对交通事故、火灾等可能突发的公共安全事件进行预判和预防；事中可以在全市甚至更大范围内实时、合理调配资源，在最短时间内将医疗、公安、消防、防疫等力量聚集到现场；事后可以借助块数据分析手段，最大限度减少伤亡，加快恢复进程。

比如，在长寿城市中，健康管理将取代疾病诊治成为个人身体管理的常态，DNA识别仪作为终端设备，将根据你的健康状况，特别是健康指标的细微差别，给予相应的预警和提示。如果你最近工作繁忙、身体有所不适，系统会自动提示你，要求你在平时的运动、饮食上注意科学搭配，针对平时工作生活状态，督促监督你调理身体、合理膳食、有序锻炼，随时对身体进行微修复，让身体处于最佳状态，不用或者很少需要因疾病去医院诊治。特别是，心理调适将成为按需定制的重要健康管理手段。通过工作数据、身体数据、生活数据、思维数据等的融合，健康管理系统能够科学确定你的压力承受范围，有针对性地进行实时心理辅导和治疗，让你尽可能地保持舒心、低压力的心理状态。

（四）惊艳旅游。根据消费者的习惯，根据城市既有的功能，对供需双方进行精准匹配，超预期安排休闲游乐活动，而且每一步都严格保密，让消费者觉得惊喜甚至是惊艳。本质上讲，惊艳旅游城市的关键

就在于给游客提供不间断的惊喜。这就要求，要把一个城市的交通、旅游、文化、气象、产业等各方面的数据全部汇聚起来，形成城市基础数据池，让城市每一个方面的特点都能够详细、准确、实时地以可视化的方式体现出来，并能够随时调取和匹配。在此基础上，要构建由旅游者身份数据、行为数据、思维数据，特别是兴趣爱好数据等关联形成的需求数据系统。这实际上就是要求尽可能大地扩充以人为识别对象的基础块数据库，通过海量条数据的交叉比对，完成对旅游者的精准数据画像，从而知道旅游者心仪的某件产品或服务，在此基础上将消费者的需求信息和城市能够提供的供给信息进行匹配，确定他近期没有得到这一产品或服务，就可以及时给他们提供各类惊喜。

比如，通过对你的历史数据、消费数据、社交数据的分析，知道你最难忘的是高中毕业前与同桌好友吃一顿烧烤、喝一瓶二锅头，但你已经与当年的好兄弟多年未见、天各一方。那当你来旅游时，块数据中心就会提醒旅游公司安排你去尝一尝特色烧烤。特别是，当块数据中心经过数据采集分析，了解到你的同桌好友正好也在这座城市，旅游公司就会安排你在不知情的情况下去偶遇好友，给你带来惊艳式体验。

（五）泛在教育。通过将每一个学生的行为数据、社交数据、健康数据、DNA数据和父母、朋友、所在社区等海量的各类条数据融合起来，综合评定学生的接受能力、兴趣爱好、学习进度和发展方向，科学制订学习方案和手段，真正实现个性化、定制化、智能化教育。与传统教育模式相比，智能化定制教育完全能够实现教育效果评价的精准化和教育方式的高度个性化。一方面，每一个学习者在学习过程中的细微变化都将被完整、全面地记录下来，再结合学习者的心理素质、家庭背景、个性特点等数据，经过融合分析，形成高度个性化的学习方案和学习进度督导机制，实现因材施教、精准培育。另一方面，借助DNA识别仪等终端设备，学习场景将与工作、生活场景链接、融合，学习者将能够在任何场景、任何时间进行有针对性的学习，问题导向的学习模式、终身学习的理念将真正成为现实。

可以预见，在定制化教育场景中，将来的学校就是一个大型的娱乐场所，学校通过组织大量的聚会、比赛、素质拓展等社交、娱乐活动，借助块数据分析手段，在潜移默化中以一种精细化、可预测的方式完成对你性格特点、家庭背景、智力水平的画像，然后有针对性地提出教育方案，真正做到因材施教。

（六）**360°全息媒体。**借助块数据手段以及深度学习技术，通过机器人的“眼睛”来看世界，撰写新闻稿件，完整、真实、客观地异地呈现新闻的全场景。媒体的发展大体经历了三个阶段，第一阶段是无组织传播，主要是在工业文明以前，由于缺乏专业化的新闻机构，新闻传播呈现松散化、小范围等特点。第二阶段是有组织传播，主要是进入工业化时代以后，新闻舆论呈现出组织化、中心化、快速化特征。第三阶段是自媒体时代，进入互联网时代，借助网络平台，人人都可以发声，每个人都有机会成为新闻传播中心。现在，随着各类数据的进一步融合、各类先进算法的进一步改进，特别是机器学习技术的进一步发展，我们正在进入360°全息媒体时代。与过去的媒体相比，360°全息媒体将实现前所未有的客观性和精准性，因为机器没有价值立场和利益诉求，借助块数据，它们知道社会的热点在哪儿，知道社会关注的点在哪儿，比一般的通讯社更了解社会需求，能够随时抓取各种社会现象进行报道。特别是，通过连接整个网络世界的实时新闻数据库，并和各类地理信息数据、公民法人数据、历史档案数据、个人生活数据交互，就可以实现新闻撰写的准确性、实时性和丰富性，甚至做到贴身式的新闻提供、深度挖掘。同时，借助虚拟现实（VR）等技术手段和块数据支撑，将实现对新闻报道的三维动态视景和事件过程的系统再现，让市民能够真实感受新闻全过程。

可以预想，借助“水晶球”媒体，我们将看到越来越多的和以往立场完全不一样的新闻报道，比如，针对某些国家的持续动荡，机器人可能会呈现大量翔实的数据和案例，但不进行价值判断，不进行善恶分析，留给读者自行研判。而不是像现在我们看到的那样，报道集中在由于独裁统治、历史传统、本身国民素质较低等原因造成的今日乱局。

（七）**消费4.0。**消费4.0是与工业4.0相对应的概念。工业4.0是指利用物联信息系统（CPS）将生产中的供应、制造、销售信息数据化、智慧化，最后实现快速、有效、个人化的产品供应，简单讲就是工业品生产的智能化、定制化、小批量化。而消费4.0则是依托块数据，从消费者角度出发，借助消费者人联网、互联网系统，完成消费的数据化、智能化和预判化，进而提供更好的产品和服务。一方面，根据消费者个人特征数据、行为数据、社交数据、网购数据、信用数据以及线上线下商铺数据等各类数据，可以了解消费者的行为习惯和偏好，从需求出发，精准提供满足消费者需求的各类产品和服务。另一方面，基于块数据融合分析，对流行趋势、消费倾向、行业走向进行预测预判，对个

人的消费内容进行智能化安排，从挖掘需求向创新供给转变，必要的时候可以根据分析结果创造供给，实现对消费者的超预期满足。

比如，基于对消费者需求的总体分析，发现现在市场上所有医疗服务APP（应用程序）均存在资料库不够丰富、反应速度过慢、缺乏疑难杂症案例等问题，然后决定打造一款名为"我的医生"的全新APP，实现智能化健康管理和疾病诊断。这就像乔布斯团队研发苹果手机一样，不是简单地停留在改善现有的产品和服务上，而是基于对市场趋势的总体把握，从0到1创造借助APP来诊断和管理的全新模式，向市场供给超预期服务，主动引领市场潮流。

（八）超级链接组织。依托财务数据、社交数据、身份组织数据、社会行为数据、工作学习数据等的汇聚分析，将人、社会组织数据化，将人与人、组织与人之间的特长、需求进行精准的匹配和链接。超链接组织实际上是一个人力资源管理的大平台，具有如下两大功能。一个是汇集功能，将有人力资源需求的个体、组织的历史记录、具体需求信息、自身发展情况、人力及行为数据进行存储和画像，对求职者的求职数据、教育数据、行为数据、社交数据等相关条数据进行抓取和分析。另一个是精准匹配功能，通过运行算法和模型，将各类企业、组织对人力资源的需求和市场上不同类型人力资源的特点进行配对，最终实现供需双方的高度匹配。

比如某人个性开朗，具有海外教育背景和工作经验，个人诚信记录良好，长期居住在本地，喜欢画画、打篮球，而某家金融类企业也在寻求开拓当地市场的营销类高管。这样，供需双方就很容相互匹配，实现人力资源的最优配置。更有趣的是，借助这个系统，也可以对婚姻恋爱等社交需求进行牵线搭桥，在双方初次见面之前就已经对匹配度进行了分析，对诚信情况有所了解，对彼此的生活习惯、学历能力背景都有所掌握，这就为双方的进一步交往提供了很好的基础条件，极大提升成功找到另一半的概率。

参考文献

[1] 中国共产党第十八届中央委员会.中国共产党第十八届中央委员会第五次全体会议公报[R] .新华网，2015-10-29.

[2] 中华人民共和国国务院.关于印发促进大数据发展行动纲要的通知[R] .中国政府网，2015-9-5.

[3] 中华人民共和国国务院.中华人民共和国国民经济和社会发展第十三个五年规划纲要[R] .新华网，2016-03-17.

[4] 马克思，恩格斯.马克思恩格斯选集（第2卷）[M] .北京：人民出版社，1995.

[5] [美] 杰里米·里夫金.第三次工业革命：新经济模式如何改变世界[M] .张体伟、孙豫宁，译.北京：中信出版社，2012.

[6] [美] 哈罗德·D.拉斯韦尔，亚伯拉罕·卡普兰.权力与社会：一项政治学研究的框架[M] .王菲易，译.上海：上海人民出版社，2012.

[7] [美] 斯蒂芬·P·罗宾斯.组织行为学[M] .孙健敏等，译.北京：中国人民大学出版社，1997.

[8] [美] 尼古拉斯·卡尔.浅薄：你是互联网的奴隶还是主宰者[M] .刘纯毅，译.北京：中信出版社，2015.

[9] [美] 奥尔霍斯特.大数据分析：点“数”成金[M] .王伟军，译.北京：人民邮电出版社，2013.

[10] [美] 凯文·凯利.必然[M] .周峰，董理，金阳，译.北京：电子工业出版社，2016.

[11] [美] 维纳.人有人的用处[M] .陈步，译.北京：商务印书馆，1978.

[12] [美] 埃尔伍德·霍尔顿.在组织中高效学习：如何把学习成果转化为工作绩效[M] .沈亚萍等，译.北京：机械工业出版社，2016.

[13] [美] W·理查德·斯科特，杰拉尔德·F·戴维斯.组织理论：理性、自然与开放系统的视角[M] .高俊山，译.北京：中国人民大学出版社，2011.

[14] [美] 乔尔·古林.开放数据[M] .张尚轩，译.北京：中信出版社，2015.

[15] [美] 乔治·索罗斯.金融炼金术[M] .孙忠，译.海口：海南出版社，1999.

[16] [美] 艾尔·巴比.社会学研究方法[M] .邱泽奇，译.北京：华夏出版社，2009.

[17] [美] 梅兰妮·斯万.区块链新经济蓝图及导读[M] .韩锋，译.北京：新星出版社，2016.

[18] [美] 纳西姆·尼古拉斯·塔勒布.黑天鹅：如何应对不可预知的未来[M] .万丹，刘宁，译.北京：中信出版社，2011.

[19] [美] 雷切尔·博茨曼，路·罗杰斯.共享经济时代[M] .唐朝文，译.上海：上海交通大学出版社，2015.

[20] [美] 马克·布尔金.信息论[M] .王恒君，嵇立安，王宏勇，译.北京：知识产权出版社，2015.

[21] [美] 托马斯·库恩.科学革命的结构[M] .金吾伦，胡新和，译.北京：北京大学出版社，2003.

[22] [美] 杰夫·霍金斯，桑德拉·布拉克斯莉.智能时代[M] .李蓝，刘知远，译.北京：中国华侨出版社，2014.

[23] [美] 杰里米·里夫金.零边际成本社会[M] .赛迪研究院专家组，译.北京：中信出版社，2014.

[24] [美] 罗宾·蔡斯.共享经济：重构未来商业新模式[M] .王芮，译.杭州：浙江人民出版社，2015.

[25] [美] 德尼·古莱.残酷的选择：发展理念与伦理价值[M] 高铦，高戈，译.北京：社会科学文献出版社，2008.

[26] [英] 维克托·迈尔-舍恩伯格.删除[M] .袁杰，译.杭州：浙江人民出版社，2013.

[27] [英] 维克托·迈尔-舍恩伯格，肯尼思·库克耶.大数据时代[M] .盛杨燕，周涛，译.杭州：浙江人民出版社，2013.

[28] [英] 亚当·斯密.道德情操论[M] .樊冰，译.太原：山西经济出版社，2010.

[29] [法] 孟德斯鸠.论法的精神（上册）[M] .严复，译.北京：商务印书馆，1982.

[30] [法] 常博逸，刘文波.轻足迹管理[M] .李瑶光，译.北京：中信出版社，2014.

[31] [法] 米歇尔·福柯.规训与惩罚[M] .刘北成，杨远婴，译.上海：上海三联出版社，2013.

[32] [日] 大栗博司.超弦理论：探究时间、空间及宇宙的本源[M] .逸宁，译.北京：人民邮电出版社，2015.

[33] [加] 萨利姆·伊斯梅尔，[美] 迈克尔·马隆.指数型组织：打造独角兽公司的11个最强属性[M] .苏健，译.杭州：浙江人民出版社，2015.

[34] [匈] 欧文·拉兹洛.自我实现的宇宙：科学与人类意识的阿卡莎革命[M] .杨富斌，译.杭州：浙江人民出版社，2015.

[35] 大数据战略重点实验室.块数据：大数据时代真正到来的标志[M] .北京：中信出版社，2015.

[36] 大数据战略重点实验室著，连玉明主编. DT时代：从“互联网＋”到“大数据×”[M] .北京：中信出版社，2015.

[37] 本书编写组.大数据领导干部读本[M] .北京：人民出版社，2015.

[38] 涂子沛.大数据[M] .桂林：广西师范大学出版社，2013.

[39] 涂子沛.数据之巅：大数据革命，历史、现实与未来[M] .北京：中信出版社，2014.

[40] 徐晋.大数据经济学[M] .上海：上海交通大学出版社，2014.

[41] 李军.大数据：从海量到精准[M] .北京：清华大学出版社，2015.

[42] 陈琳琳，李建林.数据结构与算法C语言版[M] .北京：清华大学出版社，2015.

[43] 王汉华，刘兴亮，张小平.智能爆炸[M] .北京：机械工业出版社，2015.

[44] 金观涛，刘青峰.兴盛与危机：论中国社会超稳定结构[M] .北京：法律出版社，2011.

[45] 刘大椿.自然辩证法[M] .北京：中国人民大学出版社，2006.

[46] 陈潭.大数据时代的国家治理[M] .北京：中国社会科学出版社，2015.

[47] 蔡余杰，黄禄金.共享经济[M] .北京：企业管理出版社，2015.

[48] 李耀东，李钧.互联网金融框架与实践[M] .北京：电子工业出版社，2014.

[49] 李勇，徐荣.大数据金融[M] .北京：电子工业出版社，2016.

[50] 李国华，吴博.共享经济2. 0：个人、商业与社会的颠覆性变革[M] .北京：企业管理出版社，2015.

[51] 郭庆光.传播学教程[M] .北京：中国人民大学出版社，1999.

[52] 胡泳.众声喧哗[M] .桂林：广西师范大学出版社，2008.

[53] 迪莉娅.大数据环境下政府数据开放研究[M] .北京：知识产权出版社，2014.

[54] 张民安.美国当代隐私权研究[M] .广州：中山大学出版社，2013.

[55] 张民安.隐私权的比较研究[M] .广州：中山大学出版社，2013.

[56] 王利明，杨立新等.民法学（第四版）[M] .北京：法律出版社，2015.

[57] 中国电信智慧城市研究组.智慧城市之路科学治理与城市个性[M] .北京：电子工业出版社，2013.

[58] 植草益，朱绍文.微观规制经济学[M] .北京：中国发展出版社，1992.

[59] 王忠.大数据时代个人数据隐私规制[M] .北京：社会科学文献出版社，2014.

[60] 穆胜.云组织[M] .北京：电子工业出版社，2015.

[61] 陈春花.激活个体[M] .北京：机械工业出版社，2015.

[62] 李檬.去中心化：移动互联时代电商运营制胜法则[M] .北京：人民邮电出版社，2015.

[63] 彭剑锋，云鹏.海尔能否重生：人与组织关系的颠覆与重构[M] .杭州：浙江大学出版社，2015.

[64] 万建华.金融E时代[M] .北京：中信出版社，2013.

[65] 车品觉.决战大数据[M] .杭州：浙江人民出版社，2014.

[66] 陈刚.块数据的理论创新与实践探索[J] .中国科技论坛，2015（4）.

[67] 张枝令.结构化数据及非结构化数据的分类方法[J] .宁德师专学报，2007（4）.

[68] 周宇峰.暗物质属性与探测研究进展[J] .中国科学：物理学力学天文学，2015（4）.

[69] 赵言舟，赵磊.知识是思维活动的基本构件[J] .政工学刊，1998（1）.

[70] 李天柱，王圣慧，马佳.基于概念置换的大数据定义研究[J] .科技管理研究，2015（12）.

[71] 齐磊磊.大数据经验主义：如何看待理论、因果与规律[J] .哲学动态，2015（7）.

[72] 赵毅衡.重新定义符号与符号学[J] .国际新闻界，2013（6）.

[73] 孟增辉.知识定义及转化研究[J] .计算机工程与应用，2015（13）.

[74] 刘红.大数据的本体论探讨[J] .自然辩证法通讯，2014（6）.

[75] 黄俊.杰像人脑一样感知[J] .第一财经周刊，2013（9）.

[76] 胡瑞敏，徐正金.人工神经元网络的智能神经元模型[J] .电子学报，1996（4）.

[77] 肖锋.论信息技术时代的三大认识论悖论[J] .创新，2016（1）.

[78] 高书国.大数据时代的数据困惑——教育讲究的数据困境[J] .教育科学研究，2015（1）.

[79] 刘红，胡新和.数据革命——从数到大数据的历史考察[J] .自然辩证法通讯，2013（6）.

[80] 黄欣荣.大数据哲学研究的背景、现状与路径[J] .哲学动态，2015（7）.

[81] 尹启天，马立丽.教学过程当中的混沌性[J] .今日科苑，2008（16）.

[82] 吉志伟.混沌理论在建筑设计中的运用[J] .中外建筑，2011（6）.

[83] 业茂.蝴蝶扇动了翅膀[J] .科学24小时，2005（2）.

[84] 汪斌.基于经营战略形态和产业价值链的经营战略矩阵设计探讨[J] .企业活力，2012（9）.

[85] 王文清.科学教育中的建模理论[J] .科技信息，2011（3）.

[86] 王婧韫.数据迁移的一般原则[J] .电脑开发与应用，2000（4）.

[87] 张晓青.国际人口迁移理论述评[J] .人口学刊，2001（3）.

[88] 牟焕森.协同消费的商业模式创新研究[J] .探求，2013（1）.

[89] 王建武.用SMART原则指导学习型组织共同愿景建设研究[J] .科教文汇，2010（7）.

[90] 孙玉霞，张筱薏.消费主义背景下生产与消费关系的历史透视[J] .学术研究，2008（6）.

[91] 朱福惠，王建学.论弱势群体的社会权[J] .云南大学学报，2012（6）.

[92] 高健，秦龙.论我国弱势群体的机会公平保障问题[J] .中州学刊，2014（2）.

[93] 王军.马克思经济学与社会经济学的比较研究[J] .经济纵横，2011（1）.

[94] 郑杭生.抓住社会资源和机会公平配置这个关键——党的十八大报告社会建设论述解读[J] .求是，2013（7）.

[95] 徐晓璐.“圆形监狱”大数据时代的隐忧[J] .青年记者，2014（1）.

[96] 姚相振，周睿康，范科峰.网络安全标准体系研究[J] .信息安全与通信保密，2015（6）.

[97] 赵战生.国内外信息安全标准化建设现状与发展趋势[J] .中国信息安全，2015（5）.

[98] 冯登国等.大数据安全与隐私保护[J] .计算机学报，2014（1）.

[99] 周水庚等.面向数据库应用的隐私保护研究综述[J] .计算机学报，2009（5）.

[100] 王航伟.自主安全防护技术在数据库中的应用[J] .电脑知识与技术，2015（5）.

[101] 墨菲.哪些设备最易受到攻击[J] .计算机与网络，2015（24）.

[102] 杨海澜.关于计算机防火墙安全屏障安全可靠网络防范途径思考[J] .电脑知识与技术，2016（2）.

[103] 张明远.计算机网络常见攻击手段和防范措施研究[J] .电子技术与软件工程，2016（2）.

[104] 刘念，余星火，张建华.网络协同攻击：乌克兰停电事件的推演与启示[J] .电力系统自动化，2016（6）.

[105] 宋晨光等.基于智能手表运动传感器的新型攻击及其防范[J] .通信学报，2015（S1）.

[106] 李伟明等.网络协议的自动化模糊测试漏洞挖掘方法[J] .计算机学报，2011（2）.

[107] 杨琪，龚南宁.我国大数据交易的主要问题及建议[J] .大数据，2015（9）.

[108] 王娜.打破平衡——组织优化提升效能策略[J] .科技视界，2013（24）.

[109] 宋哲，王树恩.共享型组织模式：从组织层面改善界面管理[J] .现代管理科学，2008（10）.

[110] 王成文.数据力："大数据"KO"小数据"[J] .中国传媒科技，2013（19）.

[111] 沈婷婷.数据素养及其对科学数据管理的影响[J] .数字技术，2015（1）.

[112] 张建云，毛文龙."经济人"与"社会人"的逻辑关系及当代意义[J] .经济研究导刊，2009（6）.

[113] 袁斐，朱婧.组织结构变革的动因、影响因素及途径[J] .现代

企业，2008（8）.

[114] 许宁.阿里巴巴商业模式“三级跳”：马云C2B猜想[J] .经理人，2012（9）.

[115] 姚文建.自组织理论下的国家开放大学办学体系建设探索[J] .中国远程教育，2013（5）.

[116] 张瑞敏. VUCA下的零距离和从零开始——《轻足迹管理》的三点启示[J] .商学院，2014（12）.

[117] 李红梅.数据处理能力的内涵[J] .课程教材教学研究，2014（Z6）.

[118] 郑文范，崔明浩.创客模式与战略变革契合及实现路径研究[J] .科技进步与对策，2015（17）.

[119] 许正权.组织智力与复杂性科学[J] .科学与管理，2003（1）.

[120] 成思危.复杂科学与组织管理[J] .科学，2001（1）.

[121] 崔景贵.复杂性科学视野中的学校心理教育[J] .教育导刊，2004（Z1）.

[122] 范国睿.复杂科学与教育组织管理研究[J] .教育研究，2004（2）.

[123] 高文杰.复杂性科学视域下的高职教育系统认知与管理策略[J] .中国职业技术教育，2011（27）.

[124] 李景平，刘军海.复杂科学的研究对象：非线性复杂系统[J] .系统辩证学学报，2005（3）.

[125] 苗振青，李良贤.基于系统演化理论的企业集聚式共生发展研究[J] .学术论坛，2012（6）.

[126] 秦书生.复杂性的哲学评析[J] .哲学动态，2004（5）.

[127] 金祥荣，张利风.复杂性科学与复杂性经济学[J] .经济学动态，2003（12）.

[128] 陶倩，徐福缘.基于机制的复杂适应系统建模[J] .计算机应用研究，2008（5）.

[129] 曲泽静等.新常态下价值链升级的创新驱动系统研究[J] .技术经济与管理研究，2016（1）.

[130] 胡税根等.基于大数据的智慧公共决策特征研究[J] .浙江大学学报，2015（3）.

[131] 唐斯斯等.以“数据治理”推动政府治理创新[J] .中国发展观察，2014（5）.

[132] 李玉剑，宣国良.企业供应链的价值网管理模式与现实构建[J] .经济管理，2004（8）.

[133] 李长玲，邵景.信息与知识价值链[J] .图书与情报，2004（3）.

[134] 胡雄伟，张宝林等.大数据研究与应用综述（上）[J] .标准科学，2013（9）.

[135] 高薇华.由价值链到价值网：动漫产业的内生增长模型[J] .现代传播（中国传媒大学学报），2013（8）.

[136] 林玉妹，林善浪.我国产业转型升级的关键因素与路径分析[J] .北华大学学报（社会科学版），2013（1）.

[137] 赵伶俐.量化世界观与方法论——《大数据时代》点赞与批评[D] .重庆：西南大学，2013.

[138] 陈玉. D·H劳伦斯对未知的探索[D] .成都：四川大学，2004.

[139] 刘成波.公共场所视频监控的法律规制——以隐私权保护为视角[D] .长春：吉林大学，2013.

[140] 赵阳.大数据时代对国家安全的挑战及对策研究[D] .济南：山东师范大学，2015.

[141] 曹莉兰.基于防火墙技术的网络安全机制研究[D] .成都：电子科技大学，2007.

[142] 李建军.基于逻辑的网络安全漏洞分析研究[D] .郑州：解放军信息工程大学，2012.

[143] 刘东亮.我国数据库产业发展与战略研究[D] .郑州：解放军信息工程大学，2012.

[144] 栾亚丽.民主价值论[D] .北京：中共中央党校，2007.

[145] 张兵.哲学视野下管理学中的人性假设[D] .苏州：苏州大学，2006.

[146] 万里鹏.非结构化到结构化数据转换的研究与实现[D] .成都：西南交通大学，2013.

[147] 张芳.解读《社会学想象力》[D] .北京：北京工业大学，2014.

[148] 彭立志.基于数据引力的分类方法及网络异常检测模型的研究[D] .济南：济南大学，2006.

[149] 张玉洁.综合孵化系统的复杂性仿真模拟及其网络化虚拟化研究[D] .天津：天津大学，2012.

[150] 赵生辉，汤志伟.政府电子化公共服务需求分析模型构建研究[D] .成都：电子科技大学，2007.

[151] 袁青燕.价值网的竞争优势形成机理研究[D] .南昌：江西财经大学，2013.

[152] 中央经济工作会议要求准确把握经济发展9方面新常态[EB/OL] .人民网. 2014-12-11.

[153] 贺宝成.大数据与国家治理[N] .光明日报，2014-3-27.

[154] 黄欣荣.大数据时代的哲学变革[N] .光明日报，2014-12-13.

[155] 蔡报永.点亮“暗数据”的五种方式[N] .中国计算机报，2015-1-12.

[156] 张茉楠.大数据国家战略推动“数据驱动经济”[N] .南方都市

报，2015-11-06 .

[157] 张静.专家解读“块数据”：从解构到重构，从多维到共享[EB/OL] .中国科技网，2015-4-20.

[158] 赵昌文.供给侧是一场问题导向的改革[N] .中华工商时报，2016-1-28.

[159] 赵勇.关注大数据就是关注活性与流动性[EB/OL] .财经网，2015-11-19.

[160] 姜海东.“互联网＋”的本质是数据的流动，而数据就是能源[EB/OL] .虎嗅网，2015-6-14.

[161] 干江沄，杨丹等. 2016数博会“痛客计划”新闻发布会[EB/OL] .中国大数据产业观察网，2016-3-1.

[162] 黎斌.《易经》“数相”与“大数据”[EB/OL] .中国大数据产业观察网，2016-3-18.

[163] 王长华.司法中立与权力边界[EB/OL] .共识网，2014-9-23.

[164] 张小彦.数据融合：大数据分析的瓶颈[EB/OL] .财新网，2016-3-23.

[165] 佚名.从百度技术救灾看企业践行社会责任的新模式[EB/OL] .赛迪网，2014-8-6.

[166] 解读新国家安全法：首次以法律形式提出“维护国家网络空间主权”[EB/OL] .央广网，2015-7-2.

[167] Hestenes, D. (1992). Modeling games in the Newtonian world.*American journal of Physics*, 60(60):732-748. Tempe, U. S. : Physics Department Arizona State University.

[168] Halloun, I. A. (1998). Schematic concepts for schematic modeling of real world: The Newtonian concept of force.*Science Education*, 82(2), 239-263. U. S. : National Association for Research in Science Teaching. DOI: 10. 1002/sce. 3730380217Science.

[169] Wilson, E. O. (1976). Sociology: The new synthesis.*Journal of Animal Ecology*, 46(3):28-43. Cambridge, U. K. : Harvard University Press. DOI: 10. 2307/1297251

[170] Eggers W. D. (2005). Government 2. 0: Using technology to improve education , cut red tape , reduce gridlock and enhance democracy.*Future Survey*, 310. Lanham, U. S. : Rowman & Littlefi eld Publishers. DOI: 10. 1007/978-1-4614-1448-3 16_

[171] Battiston S. J. , & Farmer D. (2016, Feb 19). Complex systems complexity theory and financial regulation: Economic policy needs interdisciplinary network analysis and behavioral modeling.*INSIGHTS: PERSPECTIVES*, 351(6275). AAAS. Retrieved Februry 26, 2016, from 10. 1126/science. aad0299.

术语索引

A

安全漏洞 43

暗数据 37

暗数据的剔除 56

暗物质 50

B

变量 7

变异和突变 133

C

参数 117

产消合一 206

产消联盟 207

超稳定结构 184

抽象化数据 37

创客运动 214

创造者经济 206

刺激响应机制 130

D

大数据× 103

大数据产业聚集试验 233

大数据产业生态体系 233

大数据创新应用试验 233

大数据国际合作试验 234

大数据时代的权力观 253

大数据制度创新试验 234

大数据资源流通试验 234

电脑时代 7

动源状态 124

多维变量 77

多维度数据分析 30

多维性 19

F

反腐防火墙 251

反腐数据库 252

反馈 62

泛金融化 219

非耗散性 128

非结构化数据 23

分层组合机制 161

分析数据集 100

服务流 3

腐败系数 258

负熵 189

复杂科学 34

复杂理论 83

复杂适应系统 115

G

干扰数据 57

公共服务协调平台 161

公权治理力 252

公众需求分类 156

共享经济 4

共享社会 4

共享型组织 4

关键信息基础设施 269

关联关系 63

关联规则 65

关联融合 122

关联系数 92

关联性集聚 61

过滤筛选机制 130

H

海量数据的悖论 28

黑客效应 282

回溯型数据决策 167

混沌序 189

J

机制 3

激活数据学 2

激活数据学扩散规律 119

激活数据学模型 128

激活效应 62

极致分工 208

极致生产力 201

棘轮效应 237

技术流 3

加法效应与乘数效应 104

价值关联 7

价值活动 141

价值链网络 147

价值生态圈 190

价值数据流 167

价值网理论 144

价值网络 118

价值转移 216

交互式协作 213

结构化数据 21

金融脱媒 220

进化机制 115

经济新动能 212

经验抽象化 49

精准供给机制 162

静态数据 70

聚合 7

聚合力指数 88

K

开放数据 5

开放政府 242

KANO模型 156

科学范式转移 80

跨界融合 4

块时代数据观 34

块数据 1

块数据安全防御 286
块数据的高度关联性 72
块数据的关联度 92
块数据的聚合力 93
块数据的开放性 92
块数据的平台化 4
块数据的社会学范式 60
块数据的不确定性 73
块数据的超时空 73
块数据的非线性 73
块数据服务 130
块数据风险源 280
块数据价值链 3
块数据价值链理论 145
块数据价值链模型 146
块数据价值中枢 139
块数据金融 215
块数据经济 2
块数据模型 87
块数据平台 94
块数据强活性 74

块数据全产业链 148

块数据全产业链体系 150

块数据全服务链 156

块数据全服务链体系 158

块数据全治理链 162

块数据全治理链体系 164

块数据主义 178

块数据组织 2

扩散 18

L

垃圾数据 3

离散化解构 25

利他主义 6

连贯机制 74

临界点 56

灵敏度 162

零边际成本 181

M

锚定因子 135

敏感数据安全 288

明数据 52

模块化供给 139

N

扭曲数据 57

P

平台化率 88

平台领导力 3

平台生态圈 98

普惠金融 220

Q

契约式开放 5

潜动源状态 124

轻资产 208

区块链 215

去中心化 184

权力的度量 255

权力数据化 4

全球价值链理论 145

全息化重构 25

全息数据 54

R

热点减量化 123

热点逻辑计算 32

热点数据 32

热点数据计算 124

热动源状态 124

人才流 3

人格数据化 166

人工智能 3

人脸识别 47

人机交互 3

人脑感知 44

人脑时代 7

容灾备份 233

冗余剔除 98

S

沙堆实验 116

社会经济 4

社会企业 227

社群经济 214

生产者经济 205

试错、容错、纠错 236

识别标志 130

适应性主体 128

数据“大同” 262

数据安全 5

数据安全保密防护体系 293

数据安全标准体系 267

数据安全观 281

数据安全监测预警系统 291

数据安全能力 294

数据安全评估体系 291

数据爆炸 29

数据壁垒 5

数据标准体系 173

“数据产生数据” 107

数据产业关联层 150

数据产业衍生层 150

数据场域 94

数据处理能力 30

数据单元 18

数据单元的状态切换 124

数据道德 301

数据道德体系 301

数据点 32

数据反腐 251

数据风险 101

数据福利 249

数据福祉 239

数据感知 139

数据共存 170

数据共享 169

数据孤岛 34

数据关系 2

数据核心层 150

数据的互利共生 103

数据化 4

数据化社会 95

数据恢复 296

数据惠民 239

数据混沌理论 63

数据混沌系统 64

数据激活能力 197

数据集 22

数据集聚 34

数据寄生 102

数据价值 41

数据价值密度 108

数据价值释放 123

数据价值挖掘 121

数据价值再造 93

数据监管 234

数据焦虑 90

数据结构化 38

数据经济 145

数据聚合 7

数据开放 5

数据空间 18

数据恐怖主义 283

数据离散化 49

数据力 2

数据立法 267

数据连接 71

数据连接力 105

数据连接网络 94

数据涟漪 104

数据流 3
数据流动 87
数据垄断 34
数据伦理 299
“数据民主”249
数据匿名 300
数据偏利共生 103
数据平衡 93
数据平衡力 106
数据平衡系统 106
数据迁移 88
数据迁移的不平衡机制 96
数据迁移的趋同机制 97
数据迁移的逐利机制 96
数据清洗 43
数据驱动 26
数据驱动性创新试验 231
数据驱动战略 149
数据去边界 102
数据权 236
数据权力化 251

数据群聚效应 102

数据群落 102

数据热点 32

数据人格 301

数据人假设 3

数据人性 301

数据融合 100

数据社会学范式 60

数据生产力 195

数据生态系统 98

数据搜索 56

“数据铁笼” 4

数据透明化 273

数据网 105

数据文化 6

数据无边界 273

数据显性连接与隐性连接 93

数据信用体系 217

数据引力 7

数据引力波 2

数据引力场 19

数据映射 51

“数据友好”型组织 198

数据宇宙 18

数据元 109

数据噪声 125

数据政府 239

数据质点 7

“数据治国”259

数据治理 4

数据中心整合利用试验 232

数据主权 6

数据资源 5

数据资源共享开放试验 232

数相 51

“数值想象力”178

思维范式 7

思想孵化器 230

所有权分离 203

T

他激活 18

他激活效应 74

“条时代”的数据观 33

痛点 223

痛点的SMART黄金原则 224

痛客计划 223

痛客经济 225

痛客治理 228

痛客智库 228

W

网络恐怖主义 283

网络拓扑结构 118

网络正效应与网络负效应 119

微笑曲线 152

无边界组织 175

无效数据 57

无组织的组织力量 184

物联网 206

物质流 3

X

线性价值链 151

消费者经济 205

消费资本化 207

协同消费 68

新技术 16

新经济 4

新模式 16

新业态 27

信号 44

信息爆炸 29

信息能量 14

信息数据化 14

行为数据 41

需求识别 139

Y

业务连续能力 288

“遗传算法” 133

意识数据 70

引爆点 6

优质数据 29

预测性决策 167

预测性数据决策 167

预防腐败的全治理链 252

预置性数据决策 167

源数据 96

云端服务器 300

云脑时代 7

云脑思维 7

Z

政府数据开放 5

知识超载 12

指向性集聚 60

治理数据 163

“智慧经济” 214

智能碰撞 32

智能运行 226

主体性 69

状态转换机制 130

资金流 3

资源分配机制 130

资源数据化 4

自雇型劳动者 208

自激活 18

自激活效应 74

自流程 21

自适应 21

自信用社会 217

自主搜索 121

自组织 21

自组织临界性 117

组合式创新 213

组织脆性 184

组织再平衡 198

后记

2015年12月1日，贵州省委常委、贵阳市委书记陈刚同志在会见微软亚洲研究院常务副院长马维英博士时提出了“激活数据学”，为《块数据2.0》的研究奠定了基础。激活数据学作为继块数据之后又一新观点和新假说，既是进行多维度大数据分析的方法论，也是基于复杂理论的新的数据社会学研究范式。

大数据战略重点实验室是由北京市科学技术委员会和贵阳市人民政府共建的跨学科、专业化、国际化、开放型研究平台。2015年5月，大数据战略重点实验室首次推出《块数据：大数据时代真正到来的标志》，创造性地提出“块数据”的概念。在此基础上，大数据战略重点实验室运用以人为原点的数据社会学分析方法，综合借鉴哲学、社会学、政治学、城市学、数学、统计学、物理学、化学、经济学、管理学、生物学、神经心理学、天文学等多种学科知识，推出了升级版的《块数据2.0》。在块数据的研究中，我们惊奇地发现了很多值得探寻的问题。这些问题超越概念或技术本身，它是边缘的、跨界的、融合的，但却是无限的、无穷的。比如，数据引力波的作用、块数据组织的魅力、“数据铁笼”的创新、数据力与数据关系的变化、数权法的期待等，这些都深刻地影响着社会关系的变革，代表着未来发展的方向和趋势，进而决定着人类文明的进程。

本书是根据陈刚关于激活数据学的理论框架，在《块数据》的基础上推出的又一创新成果，由大数据战略重点实验室汇聚一批大数据领域专家学者、实践者和政策研究者集中撰写。在本书的研究和写作过程中，陈刚提出总体思路和核心观点，连玉明对框架体系进行了总体设计，由朱颖慧、武建忠、张涛细化提纲和主题思想，主要由连玉明、朱颖慧、武建忠、张涛、王黔、刘春辉、陈栋、宋青、秦坚松、张俊立、王新江、宋希贤、朱盼盼、张清、贺羽、陈曦、陈盈瑾、任祉静、龙荣远、张松群、张红阳素、郝无穷、王琨撰写。北京市科委主任闫傲霜，贵阳市人民政府市长刘文新，贵州省科技厅厅长廖飞，贵州省金融办公室主任李瑶，贵州省经信委副主任康克岩，贵阳市委副书记张平，贵阳市委常委、秘书长聂雪松，贵阳市委常委、副市长刘春成，贵阳市人民政府副市长徐昊，北京大学数学科学学院教授宋春伟，中国科学院大学

管理学院教授徐艳梅，北京国家会计学院教授卢力平对本书提出了许多前瞻性和建设性意见，进一步丰富了本书的思想体系和理论体系。应该说，本书是集体智慧的结晶。

在本书写作过程中，成都数联铭品科技有限公司（BBD）创始人周涛、首席执行官曾途等人提出了大量宝贵的建议和意见。特别是周涛、曾途、陈东等多次参与块数据研讨会，以头脑风暴的方式，对块数据、痛客计划等提出了具有前瞻性、创造性的观点和理念，为本书的内容创新做出了丰富的思想贡献。此外，中信出版集团总裁王斌对本书的出版给予了大力支持，中信出版集团前沿经济社社长蒋永军组织多名编辑精心编校、精心设计，保证了本书如期出版。在此一并表示衷心的感谢！

《块数据2.0》是基于复杂理论对块数据本质及其规律的理论探索。在研究和编著本书过程中，我们尽力搜集最新文献、吸纳最新观点，以丰富本书的思想。尽管如此，由于著者水平所限，研究内容涉及众多学科领域，难免有疏漏之处，特别是对引用的文献资料及其出处也难免挂一漏万，恳请读者批评指正。

大数据战略重点实验室
2016年3月31日于北京